

MIGRAÇÃO DOS DADOS DA GENERALI PARA A TRANQUILIDADE

Nayara Mahomed Afzal

Relatório de Estágio apresentado como requisito parcial
para obtenção do grau de Mestre em Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

MIGRAÇÃO DOS DADOS DA GENERALI PARA A TRANQUILIDADE

por

Nayara Mahomed Afzal

Relatório de Estágio apresentado como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em Gestão do Conhecimento e Business Intelligence.

Orientador: Mauro Castelli

Agosto 2021

AGRADECIMENTOS

Em primeira instância , agradeço e louvo a Allah , que por sua inquestionável grandeza proporcionou-me vida, saúde , sustento, inteligência e coragem pois nada poderia alcançar sem a sua infinita bênção e misericórdia.

Aos meus queridos pais, Ahmed Afzal Mahomed e Mohshena Abdul Kader, pela educação , força, paciência e pelo amor incondicional proporcionado a mim nestes anos de vida.

Aos meus irmãos, Muhammad Nabil Afzal e Nurein Mahomed Afzal, pela companhia , amizade e pelos anos de vida que junto crescemos e vivemos.

A toda minha família , por em mim acreditar e hoje se alegram por mais uma conquista.

A esta faculdade, pela qualidade de ensino prestado e, a todos os professores Doutores dos cursos de Pós Graduação em Sistemas de Informação Empresariais e Mestrado em Gestão de Informação com Especialização em Gestão do Conhecimento e Business Intelligence.

A Tranquilidade, pela oportunidade de realizar o trabalho , à Custódia Freixial e ao colaborador Vladimir Flores e outros integrantes da equipa pela disponibilidade do tempo e recursos que contribuíram para a realização deste trabalho.

A todos que contribuíram de forma direta e indireta para este fim, o meu humilde e sincero agradecimento!

RESUMO

O desenvolvimento e a evolução exponencial das tecnologias de informação no final do século XX e no presente século XXI tornaram-se cada vez mais imprevisíveis, criando assim um dinamismo muito forte no ambiente de negócios, beneficiando quem tem acompanhado este crescimento.

Atualmente, não basta que as empresas tenham sistemas de informação que simplesmente armazenem uma grande quantidade de dados. Portanto, é necessário ter alguma capacidade de análise relevante e oportuna sobre os dados armazenados, abrindo espaço para o BI. Business Intelligence (BI), é um conjunto de ferramentas que permite a todos em uma organização o acesso de forma rápida e fácil de qualquer dado na empresa, ajudando-os a tomar decisões inteligentes e eficazes.

O principal objetivo deste estágio foi implementar e desenvolver processos de Extraction-Transformation-Loading (ETL) que efetuam grandes transformações aos dados provenientes do Koolsite -sistema Generali para o sistema Cogen -sistema Tranquilidade por forma a ter um sistema único que permitisse ser mais simples e eficientes e de fácil acesso aos utilizadores finais.

Este relatório começa apresentando o contexto e os objetivos, seguido pela descrição detalhada do contexto do problema. Em seguida, é apresentada a Revisão da Literatura seguida da metodologia e Ferramentas utilizadas. Para finalizar, é apresentado as Atividades Desenvolvidas, a Conclusão e a Bibliografia.

PALAVRAS-CHAVE

Sistemas de Informação; dados; Business Intelligence; Integração; Seguradora; Data Warehouse; ETL

ABSTRACT

The development and exponential evolution of information technologies at the end of the 20th century and in the present 21st century have become progressively unpredictable, creating this way a very strong dynamism in the business environment, benefiting those who have followed this growth.

Currently, it is not enough for companies to have information systems that simply store a large amount of data. Therefore, it is necessary to have some relevant and timely analysis capacity on the stored data, making room for BI. Business Intelligence (BI) is a set of tools that enables everyone throughout an organization to quickly and easily access any data in the enterprise helping them to make smart and effective decisions.

The goal of this Internship report was to implement and develop Extraction-Transformation-Loading (ETL) processes to integrate the data coming from the Koolsite - Generali system for the Cogen system - Tranquilidade system in order to have a single system that would allow it to be simpler efficient and easily accessible to end users.

This report starts by introducing the context and goals, followed by the detailed description of the background and the problem context. After it, is presented the Literature Review followed with the methodology and Tools used. To finalize, it is presented the Development Cycle, the Conclusion and References.

KEYWORDS

Information Systems; data; Business Intelligence; Integration; Insurance Company; Data Warehouse; ETL

ÍNDICE

1. Introdução	1
1.1. Contexto Académico	2
1.2. Contexto Empresarial	1
1.3. Objetivos Do Estágio.....	5
2. Revisão Da Literatura.....	6
2.1. Conceitos Básicos	6
2.2. Business Intelligence	6
2.2.1. Estado Da Arte.....	7
2.2.2. Arquitetura Dos Sistemas De BI	8
3. Tecnologias E Ferramentas.....	15
3.1. Tecnologias :	15
3.1.1. SQL- Structured Query Language:	15
3.1.2. PL-SQL - Procedural Language/Structured Query Language.....	15
3.2. Ferramentas:	15
3.2.1. Microsoft SQL SERVER :	15
3.2.2. Visual Studio Team System:	15
3.2.3. Pentaho Data Integration:.....	16
4. Metodologia	18
5. Trabalho DESENVOLVIDO	25
5.1. Etl's Desenvolvidos :	29
5.1.1. ETL GP_Getdatatostaging:.....	29
5.1.2. ETL GP_Staging_Controlo:.....	34
5.1.3. ETL GP_Getdatatogpf:	35
5.1.4. ETL GP_UPDATE_APOLICES_CONTROLO:	41
6. Conclusões.....	43
7. Limitações E Recomendações Para Trabalhos Futuros	44
8. Bibliografia.....	45
9. Anexos	47

ÍNDICE DE FIGURAS

Figura 1-Maiores Grupos Seguradores Europa 2019	1
Figura 2-Generali	1
Figura 3-Marcas da Generali	2
Figura 4-Maiores Grupos Seguradores de Portugal 2019	3
Figura 5-Modelo de Governo Corporativo	4
Figura 6-Área Technology and Operations	4
Figura 7-Objetivo do Estágio	5
Figura 8-Quadrante Mágico para plataformas de Analytics e BI-(Gartner 2021).....	7
Figura 9-Arquitetura de um Sistema de BI	8
Figura 10-Processo ETL.....	9
Figura 11-Abordagem de Inmon	10
Figura 12-Abordagem de Kimball.....	11
Figura 13-Exemplo Star Schema.....	12
Figura 14-Exemplo Snowflake schema	13
Figura 15-Operações em OLAP.....	14
Figura 16-Interface do Spoon.....	16
Figura 17-Exemplo de uma transformação no Spoon	17
Figura 18-Objetivos do Projeto Integração	18
Figura 19-As cinco fases de Integração	19
Figura 20- Workstreams Janeiro 2020	20
Figura 21-Workstreams Janeiro 2020 (continuação)	20
Figura 22-Workstreams Junho 2020	21
Figura 23-Macro Cronograma do Projeto de Integração	21
Figura 24-Migração Automática e Manual	23
Figura 25-Cronograma da Migração Automática.....	23
Figura 26-Modelo de Dados da Generali Portugal.....	25
Figura 27-tabelas de Apólices	27
Figura 28-Ficheiro de Recibos	28
Figura 29-SPA Portfolio	29
Figura 30-Exemplo de um cliente pesquisado no SPA	30
Figura 31-ETL GP_GetDataToStaging	31
Figura 32-Carregamento das tabelas Mestres	31
Figura 33-Carregamento dos Códigos Postais	32
Figura 34-carregamento da tabela de eventos	32

Figura 35-Paralelismo de JOBs	32
Figura 36-Tipos de Eventos	33
Figura 37-Carregamento Inicial do SPA.....	33
Figura 38-ETL GP_Staging_Controlo	34
Figura 39-Carregamento das tabelas MIG CONTROLO	34
Figura 40-Carregamento da tabela APOLICES_EXCLUSAO.....	35
Figura 41-Carregamento da tabela GPF.APOLICES_CONTROLO	35
Figura 42-Modus Operandi da migração automática	35
Figura 43-ETL GP_GetDataToGPF.....	36
Figura 44-10 Threads para a cópia dos dados.....	36
Figura 45-GetApolicies	37
Figura 46-Migração por apólice	38
Figura 47-Cópia das tabelas principais.....	38
Figura 48-Migração das tabelas	39
Figura 49-Cópia das tabelas de Apólices.....	39
Figura 50-Cópia das tabelas de Sinistros.....	40
Figura 51-Cópia das tabelas de Recibos	40
Figura 52-Cópia das tabelas de Riscos	41
Figura 53-ETL GP_UPDATE_APOLICES_CONTROLO	41
Figura 54-Transformação UPDATE_COGEN_CLAIM_W_TO_X	42
Figura 55-Transformação UPDATE_STATUS_LIBERTA_TO_X.....	42

ÍNDICE DE TABELAS

Tabela 1-Análise Preliminar das Apólices.....	22
Tabela 2-Informação do Cliente.....	25
Tabela 3-Informação das apólices.....	26
Tabela 4-Informação dos Recibos	27
Tabela 5-informação dos Sinistros	28
Tabela 6-Tempo de execução de 1milhão de apólices	37

LISTA DE SIGLAS E ABREVIATURAS

APS	Associação Portuguesa de Seguros
ASF	Autoridade de Supervisão de Seguros e Fundos de Pensões
BD	Base de Dados
BI	Business Intelligence
DM	Data Mart
DW	Data Warehouse
ETL	Extract, Transform, Load
GP	Generali Portugal
GPF	Generali Portugal Freeze
OLAP	Online Analytical Processing
PDI	Pentaho Data Integration
PMI	Project Management Institute
SA	Staging Area
SARL	Sociedade Anónima de Responsabilidade Limitada
SPA	Single Point of Access
SQL	Structured Query Language
SSAS	SQL Server Analysis Services
SSIS	SQL Server Integration Services
SSMS	SQL Server Management Studio
SSRS	SQL Server Reporting Services
SU	Seguradoras Unidas
TFS	Team Foundation Server

1. INTRODUÇÃO

Segundo o relatório sobre o Ranking dos maiores grupos seguradores europeus, realizado pela MAPFRE Economics (2020), os 15 maiores grupos seguradores aumentaram seus prémios em 5,2% face a 2018, evidenciando uma progressão média anual de 2,1% entre 2009 e 2019.

Durante a década do estudo, os grupos Axa, Allianz e Generali foram os líderes no ranking, representando 42% dos prémios de todos os grupos do ranking.

Maiores Grupos Seguradores da Europa 2019			
GRUPO	SEDE	PRÉMIOS <small>Mil Milhões de euros</small>	VARIAÇÃO 2019/18
1 AXA	França	99,9	4%
2 ALLIANZ	Alemanha	82,9	7%
3 GENERALI	Itália	69,8	5%
4 ZURICH	Suíça	42,9	8%
5 PRUDENTIAL	Reino Unido	40,3	4%
6 TALANX	alemanha	39,5	13%
7 CRÉDIT Agricole	França	37,0	10%
8 AVIVA	Reino Unido	35,6	10%
9 CNP	França	33,4	4%
10 BNP PARIBAS Cardif	França	23,9	-1%
11 MAPFRE	Espanha	23,0	2%
12 AEGON	Holanda	18,1	-6%
13 POSTE VITA	Itália	18,0	7%
14 ERGO	Alemanha	17,7	-1%
15 COVÉA	França	17,5	3%

Fonte: Mapfre Economics
Valores de 2019

Figura 1-Maiores Grupos Seguradores Europa 2019

De acordo com a figura 1, o Grupo Generali manteve a terceira posição no ranking em 2019 com 69.785 milhões de euros em prémios, obtendo um desenvolvimento positivo em ambos os segmentos de negócio 4,7% no ramo Vida e 4,5% no ramo Não Vida.

A Generali está presente em Portugal desde 1942 e opera através das duas companhias Generali Vida Companhia de Seguros, operando no ramo vida, e Generali Companhia de Seguros, ativa no ramo não vida. Em 2018, os resultados dos prémios brutos totais foram de €215 milhões (€154 milhões no ramo Não Vida e €61 milhões no ramo Vida).

Em 2019, a Generali assinou um acordo de aquisição em Portugal de 100% da companhia Seguradoras Unidas e da sociedade de serviços AdvanceCare com a Calm Eagle Holdings SARL e a Calm Eagle Parent Holdings II SARL, entidades detidas maioritariamente por fundos de investimento geridos por filiais da Apollo Global Management, pelo valor de compra de €510 milhões pela Seguradoras Unidas e €90 milhões pela AdvanceCare.

Com a aquisição da Seguradoras Unidas, a Generali conquistou uma oportunidade única de otimizar a sua posição estratégica em Portugal, tornando-se a segunda maior companhia no ramo não vida, com uma quota de mercado de 18.7%.

De acordo com Jaime Anchústegui Melgarejo, CEO Internacional da Generali, esta transação também permitiria capturar futuras oportunidades de crescimento, atingir sinergias significativas de custo, através da integração da Generali Portugal e Seguradoras Unidas, e assegurar a solidez operacional da Europ Assistance no mercado português.

Em Janeiro de 2020, o processo de fusão foi concluído com sucesso e a Seguradoras Unidas S.A. incorporou a Generali Companhia de Seguros, S.A. e a Generali Vida Companhia de Seguros S.A. tendo

alterado a sua designação para Generali Seguros S.A.. A fusão legal foi registrada comercialmente no dia um de outubro de 2020.

No âmbito desta fusão ,o presente relatório de estágio visa destacar as principais atividades realizadas e relatar a integração dos dados da Generali no sistema de Informação da Seguradora Tranquilidade-marca portuguesa que faz parte do grupo Seguradoras Unidas, S.A..

Este trabalho ,encontra-se organizado em capítulos. No primeiro capítulo apresenta-se um enquadramento do estágio e da empresa onde o trabalho apresentado foi desenvolvido , a seguradora Tranquilidade. No segundo capítulo dá-se um enquadramento teórico e uma breve abordagem de conceitos fundamentais aplicados ao longo do trabalho. No terceiro capítulo encontra-se uma descrição das tecnologias e as ferramentas utilizadas ao longo do estágio. Em seguida , o capítulo quarto descreve o desenvolvimento das atividades realizadas . Por fim, no quinto capítulo é feita uma breve análise crítica ao estágio e às tarefas desempenhadas .

1.1. CONTEXTO ACADÉMICO

O estágio foi desenvolvido no âmbito da unidade curricular – Dissertação/Projeto/Relatório de estágio do 2º ano do plano de estudos do Mestrado em Gestão da Informação com especialização em Gestão do Conhecimento e Business Intelligence, proposto pela Faculdade Nova Information Management School da Universidade Nova de Lisboa, com início em setembro e duração de um ano letivo e com a avaliação de 35 créditos.

O estágio relatado foi oferecido pela Novabase ,para trabalhar diretamente com um dos seus clientes, a seguradora Tranquilidade e teve como objetivo principal a migração de Dados da Seguradora Generali para o sistema de informação da Seguradora Tranquilidade. O estágio teve início a 22 de Setembro de 2020 e foi concluído no dia 22 de Maio de 2021, decorrido nas instalações da Tranquilidade , na Avenida da Liberdade.

A orientação pedagógica ficou atribuída ao Professor Doutor Mauro Castelli , docente da Faculdade NOVA IMS.

1.2. CONTEXTO EMPRESARIAL

O Grupo Generali foi criado no dia 26 de Dezembro em 1831 na cidade Trieste, com o nome de Imperial de Regia Privilegiata Compagnia di Assicurazioni Generali Austro-Italiche, por iniciativa do empresário Giuseppe Lazzaro Morpurgo.

Dez anos mais tarde, já estava presente nos principais centros Italianos e nos principais portos marítimos europeus. Em 1848, o grupo decidiu mudar o seu nome para Assicurazioni Generali e adotou o leão de S. Marcos como o seu símbolo. Após a Primeira Guerra Mundial (1918), Trieste foi anexada à Itália, portanto a empresa tornou-se Italiana.

Esta seguradora está presente em mais de 50 países do mundo, com quase 72.000 funcionários servindo 61 milhões de clientes.

A Generali tem uma posição de liderança na Europa e uma presença crescente na Ásia e na América Latina. A ambição da Generali é ser o parceiro para a vida dos seus clientes, oferecendo soluções inovadoras e personalizadas graças a uma rede de distribuição inigualável.



Figura 2-Generali

A holding que controlava a seguradora portuguesa Tranquilidade estava à venda, no início do ano 2019 e estava a atrair compradores espanhóis e belgas. Entre os interessados encontravam-se a Ageas, a Generali, a Zurich a Allianz e a Mapfre.

Depois do fim do prazo da entrega de propostas a melhor proposta foi a da Generali.

Assim o fundo norte-americano Apollo vendeu a Seguradoras Unidas aos italianos da Generali, numa operação que inclui as seguradoras Tranquilidade (ex-BES), Açoreana (ex-Banif) e Logo por 600 milhões de euros.

Em Janeiro de 2020, o processo de fusão foi concluído com sucesso e a Seguradoras Unidas S.A. incorporou a Generali Companhia de Seguros, S.A. e a Generali Vida Companhia de Seguros S.A. tendo alterado a sua designação para Generali Seguros S.A.. A fusão legal foi registrada comercialmente no dia um de outubro de 2020.

Entretanto também se procedeu à alteração da denominação de Seguradoras Unidas, S.A. para Generali Seguros, S.A., GENERALI, entidade única que detém a globalidade dos direitos, obrigações e posições contratuais, antes detidas pelas 3 sociedades, Generali, Generali Vida e Seguradoras Unidas.



Figura 3-Marcas da Generali

O grupo Seguradoras Unidas, S.A. é uma companhia de seguros resultante da fusão por incorporação na Companhia de Seguros Tranquilidade, S.A., das companhias de seguros T-Vida Companhia de Seguros, S.A., a Seguros LOGO, S.A. e a Açoreana Seguros, S.A..

A companhia de seguros Tranquilidade surgiu a 22 de Agosto de 1871 , no Porto ,com o nome de 'Companhia de Seguros Tranquilidade Portuense – Companhia de Seguros contra Fogo', tendo desde então vindo a acumular uma vasta experiência em todos os ramos de seguros.

Em 15 de janeiro de 2015, com a derrocada do Grupo Espírito Santo, a totalidade do capital acionista da Companhia foi adquirida pela Calm Eagle Holdings, S.ARL , uma sociedade controlada pelos fundos de investimento afiliados da Apollo Global Management, LLC, tendo esta aquisição merecido a aprovação prévia da Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF).

Em 2016 ,adquire a Açoreana Seguros, concluindo no final do ano ,o processo de fusão legal das seguradoras T-Vida, Logo e Açoreana na Tranquilidade, e alteração da sua denominação social para Seguradoras Unidas.

A fusão legal trouxe um ganho de dimensão e de escala e reforçou a oferta de produtos e serviços para Clientes .

Em 2019 , de acordo com o relatório anual da APS representado na figura 4 ,as principais alterações no capital nas companhias em Portugal sucedidas resultaram em que a Generali, integrando a Seguradoras Unidas, ultrapassa 1,1 mil milhões de euros de prémios, consolidando o terceiro lugar da adquirida no ranking com 8,9% de quota de mercado.

Maiores Grupos Seguradores em Portugal em 2019								
Nº DE ORDEM #2019	Nº2018	GRUPO SEGURADOR	SECTOR		SEGURADORAS	QUOTA DE MERCADO		CRESCIMENTO 2019 / 2018
			PRÉMIOS 2018	PRÉMIOS 2019		2018	2019	
1	1	Fidelidade	4.690.771	3.357.259	Fidelidade, Via Directa, Multicare, Fidelidade Assistance	35,7%	26,9%	-28%
2	2	ageas	2.143.047	2.107.530	Ocidental Vida, Ocidental, Ageas Não Vida, Ageas Vida, Médis	16,3%	16,9%	-2%
3	4	Seguradoras Unidas	799.844	872.357	Seguradoras Unidas (Tranquilidade, Açoreana, LOGO)	6,1%	7,0%	9%
4	6	BPI	516.988	830.836	BPI	3,9%	6,7%	61%
5	3	Santander	1.100.081	804.953	Santander Totta Vida, Aegon Vida, Aegon Não Vida, Popular	8,4%	6,5%	-27%
6	8	Zurich	438.341	770.503	Zurich Insurance, Zurich Vida	3,3%	6,2%	76%
7	7	GNB	449.622	669.664	GNB Vida, GNB Seguros	3,4%	5,4%	49%
8	5	Allianz	693.178	642.111	Allianz	5,3%	5,1%	-7%
9	9	Lusitania	414.913	467.464	Lusitania Vida, Lusitania Seguros, N Seguros	3,2%	3,7%	13%
10	11	Liberty	320.390	296.070	Liberty	2,4%	2,4%	-8%
11	10	Bankinter	320.703	266.089	Bankinter	2,4%	2,1%	-17%
12	12	Generali	218.702	232.655	Generali, Generali Vida, Europ Assistance	1,7%	1,9%	6%
13	13	CA	173.717	189.599	CA Seguros, CA Vida	1,3%	1,5%	9%
14	14	Mapfre	143.238	145.054	Mapfre Gerais, Mapfre Vida, Mapfre Asitencia	1,1%	1,2%	1%
15	15	Victoria	121.945	125.552	Victoria Seguros, Victoria Vida	0,9%	1,0%	3%
16	18	Una	81.014	118.079	Una Vida, Una Seguros	0,6%	0,9%	46%
17	16	Real	92.681	108.371	Real	0,7%	0,9%	17%
18	17	Metlife	85.801	93.950	Metlife Europe, Metlife Europe Insurance	0,7%	0,8%	9%
19	19	Caravela	59.495	76.437	Caravela	0,5%	0,6%	28%
20	20	AIG	50.769	50.119	AIG	0,4%	0,4%	-1%
21	21	Cardif	46.811	45.926	Cardif Vie, Cardif Divers	0,4%	0,4%	-2%
22	23	COSEC	37.620	39.389	COSEC	0,3%	0,3%	5%
23	22	Chubb	37.978	36.443	Chubb	0,3%	0,3%	-4%
24	24	Atradius	17.914	17.555	Atradius	0,1%	0,1%	-2%
25	26	Prévoir	12.954	15.664	Prévoir	0,1%	0,1%	21%
26	25	BBVA	15.506	15.357	BBVA	0,1%	0,1%	-1%
27	28	Abarca	8.143	15.310	Abarca	0,1%	0,1%	88%
28	27	Mútua dos Pescadores	9.105	10.665	Mútua dos Pescadores	0,1%	0,1%	17%
29	29	Coface	7.653	8.489	Coface	0,1%	0,1%	11%
30	30	Axeria	7.638	7.638	Axeria	0,1%	0,1%	0%
31	32	Hiscox	6.600	7.391	Hiscox	0,1%	0,1%	12%
32		AXA		6.774	Axa France IARD, Axa France Vie	0,0%	0,1%	
33	31	CESCE	7.266	6.378	CESCE	0,1%	0,1%	-12%
34	33	Espana	4.079	4.465	Espana	0,03%	0,04%	9%
35	34	RNA	2.938	3.636	RNA	0,02%	0,03%	24%
36		Asisa	590	2.642	Asisa Asistencia Sanitaria, Asisa Vida	0,00%	0,02%	348%
37	36	AWP	1.715	2.458	AWP	0,01%	0,02%	43%
38	35	Inter Partner	2.114	1.902	Inter Partner	0,02%	0,02%	-10%
39	38	ARAG	936	1.336	ARAG	0,01%	0,01%	43%
40	37	ACP	1.167	1.165	ACP	0,01%	0,01%	0%
41	39	Lloyd's	145	924	Lloyd's	0,00%	0,01%	
42		Planicare		98	Planicare		0,00%	
TOTAL			13.144.113	12.476.257				-5,1%
Unidade: Mil euros Fonte: APS Tratamento: ECD/Quercus								

Unidade: Mil euros
Fonte: APS
Tratamento: ECDseguros

Figura 4-Maiores Grupos Seguradores de Portugal 2019

Em 8 de janeiro de 2020, com a alteração do controlo acionista, existiram igualmente alterações ao nível da composição dos órgãos sociais.

De acordo com o Modelo de Governo Societário adotado, a Companhia integra um Conselho de Administração, um Conselho Fiscal, um Revisor Oficial de Contas e um Administrador Delegado com Funções Executivas que integra um Comité de Gestão Executivo.

Para além do Comité de Gestão Executivo, foram criados Comités Operacionais, de forma a permitir a realização, de forma permanente, da avaliação, controlo, monitorização, discussão e tomada de decisão sobre diversos aspetos relevantes do sistema de governo, sistema de gestão de risco, estratégia, objetivos e dados de negócio da Companhia.

Estes Comités, consoante a sua natureza e respetivo âmbito, são compostos por membros dos Órgãos de Administração e Fiscalização, de Titulares de Funções-Chaves e Responsáveis de Topo das diversas Direções da Companhia, permitindo por esta via uma permanente interação entre representantes das três de linhas de defesa previstas no âmbito do sistema de gestão de riscos em vigor na Companhia. Devido às alterações na Companhia e ao processo de fusão, os Comités não

funcionaram o ano inteiro.

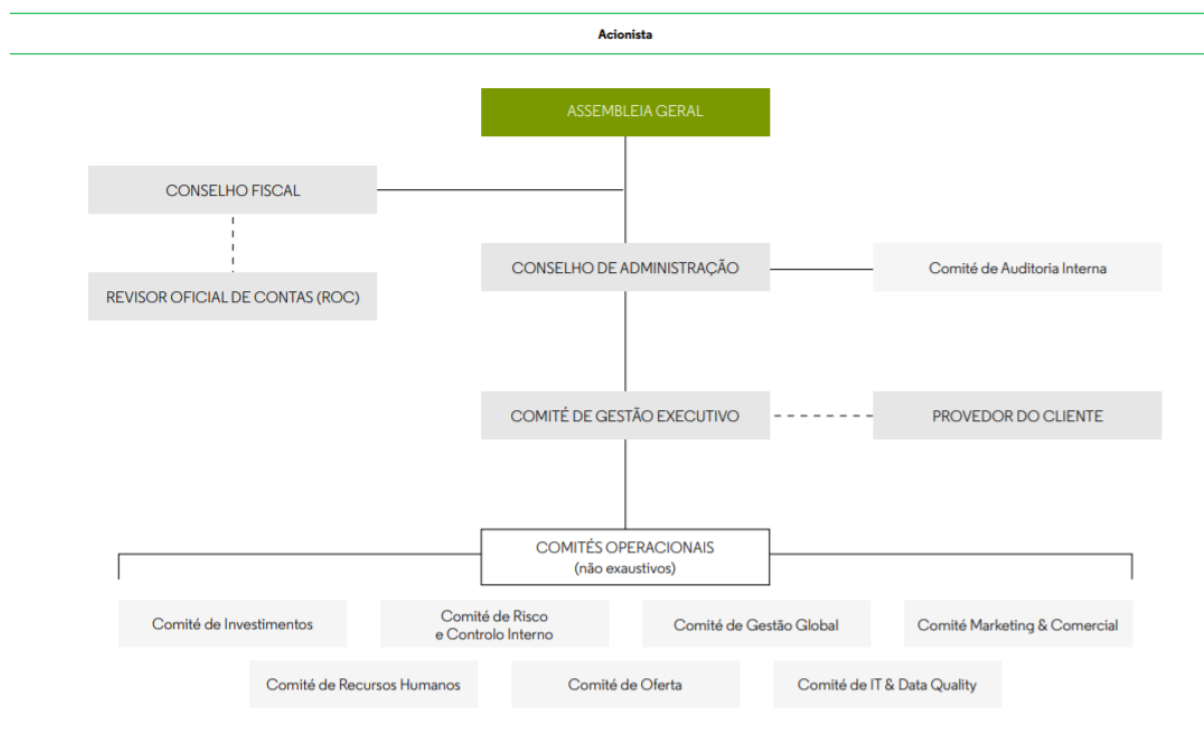


Figura 5-Modelo de Governo Corporativo

Internamente foi criada a *Technology and Operations Area* com o objetivo de garantir o pleno, eficaz e eficiente funcionamento dos processos da Companhia. Para responder a este triplo desafio, esta área foi dividida num conjunto de direções "operacionais" que lideram verticalmente um conjunto de áreas fundamentais da Companhia (*Customer Care, Architecture & Technology Infrastructure, Data, Core Development & Services, Digital Development, Operations, LOGO*).



Figura 6-Área Technology and Operations

Como estagiária foi-me dada a oportunidade de integrar no departamento Core Development & Services que tem como principal objetivo o desenvolvimento das várias aplicações e a execução da integração de sistemas.

1.3. OBJETIVOS DO ESTÁGIO

O estágio foi estruturado nas seguintes fases:

Integração : Durante as primeiras duas semanas de estágio(últimas duas semanas do mês de Setembro) o objetivo era conhecer a empresa e as ferramentas importantes a serem usadas no projeto. Na primeira semana recebi formação sobre o Pentaho – Data Integration ,uma ferramenta usada na Tranquilidade para a criação dos ETLs. Na segunda semana , realizei algumas formações do Pluralsight sobre o SQL,T-SQL e o Microsoft SQL Server Integration Service.

Projeto: No mês de Outubro, os ETLs da Integração de dados foram apresentados.

Os ETLs GP_GetDataToStaging e GP_GetDataToGPF são os dois etls principais a serem desenvolvidos durante o estágio ,responsáveis pela integração dos dados do Koolsite -sistema core da Generali Portugal no sistema da Tranquilidade-Cogen .

O objetivo principal destes ETLs é a cópia dos dados da Staging Area Gensudw- uma área intermediária que armazena os dados do Koolsite para a Staging Area Staging_GP , uma área de armazenamento intermediária que fica entre a fonte de dados e o destino de dados e é usada para processamento de dados (fase de transformação) durante o processo ETL.

Os dados do Koolsite estão armazenados numa base de dados Oracle são extraídos para uma base de dados intermediária (GENSUDW) através dum programa. A GENSUDW é uma base de dados SQL SERVER que é atualizada dinamicamente conforme as alterações que são feitas no Koolsite .Esta Base de dados contém as principais tabelas com a informação dos Clientes, apólices , recibos e sinistros bem como a informação da atividade comercial.

Após a cópia e transformações dos dados na STAGING_GP , os dados são migrados para o Sistema Origem através de programas em COBOL.

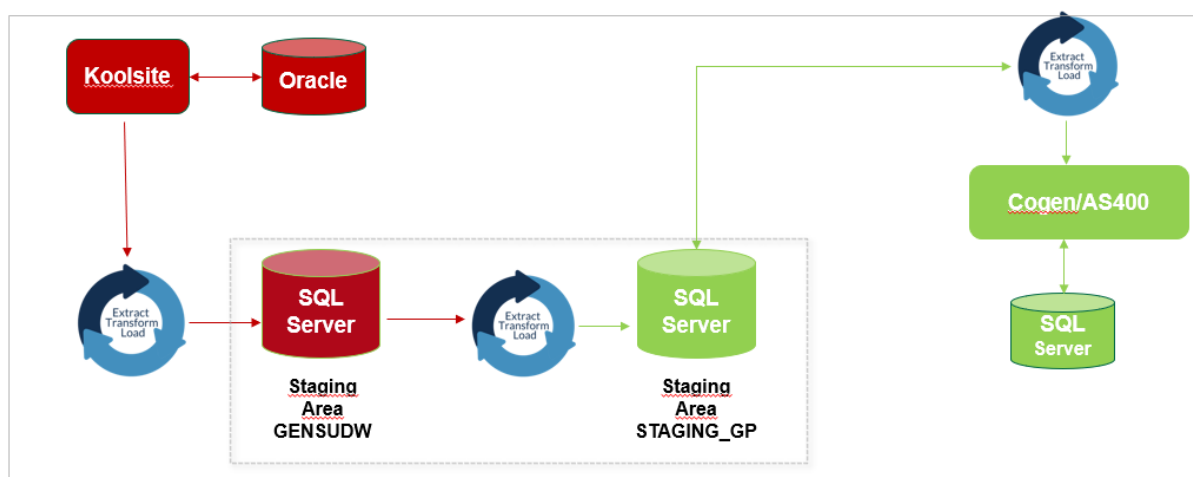


Figura 7-Objetivo do Estágio

2. REVISÃO DA LITERATURA

A revisão da literatura tem como objetivo proporcionar o enquadramento teórico ao projeto. De seguida será apresentado os conceitos técnicos mais destacados na bibliografia resgatada que estiveram relacionados com o desenvolvimento do projeto . Neste caso foram ETL e Business Intelligence, conceitos esses , que também foram explorados durante as cadeiras de Business Intelligence I e II do Mestrado.

2.1. CONCEITOS BÁSICOS

Dado, informação e conhecimento são conceitos que frequentemente estão relacionados com os sistemas de informação.

Dados são códigos que constituem a matéria-prima da informação, ou seja, é a informação não tratada. Os dados representam um ou mais significados que isoladamente não podem transmitir uma mensagem ou representar algum conhecimento (Reisswitz, 2012).

Conforme Reisswitz, (2012), informações são dados tratados. O resultado do processamento de dados são as informações, isto é, as informações têm significado. Lins (2003) reforça que a informação é gerada a partir da interpretação de um ou do conjunto de dados e que geralmente têm uma utilidade limitada.

O conhecimento vai além de informações, o conhecimento é a informação trabalhada e para além de ter um significado tem uma aplicação (Reisswitz, 2012). Ainda, Lins (2003) comenta que o conhecimento é uma informação valiosa porque está inserida num contexto, sendo por isso útil.

Uma Base de dados é um repositório que armazena dados de forma estruturada, permitindo que todos possam compartilhar, gerenciar e usar os dados.

2.2. BUSINESS INTELLIGENCE

O termo Business Intelligence foi usado pela primeira vez por Richard Millar Devens em seu livro "Cyclopaedia of Commercial and Business Anecdotes" em 1865. Neste livro é descrita a história do bancário, Sir Henry Furnese , um banqueiro de sucesso, que lucrou a partir de informações, reunindo ativamente e agindo com base nelas antes de sua concorrência, que é o conceito puro do BI.

Durante a última década de 1800, Frederick Taylor introduziu o primeiro sistema formalizado de análise de negócios nos Estados Unidos. O seu sistema de gestão científica começou com estudos de tempo que analisavam técnicas de produção e movimentos corporais dos trabalhadores para encontrar maiores eficiências que impulsionassem a produção industrial.

Taylor acabou se tornando um consultor de Henry Ford , que no início dos anos 1900 começou a medir o tempo que cada componente de seu Ford Modelo T demorava para ser concluído na sua linha de montagem. O sucesso do seu trabalho revolucionou a indústria em todo o mundo.

Em 1958 ,Hans Peter Luhn, pesquisador da IBM, escreveu um artigo onde teorizou sobre o potencial de um sistema de “disseminação seletiva” de documentos para “pontos de ação” com base em “perfis de interesse”. Seu trabalho tem um significado notável até os dias de hoje, já que ele previu várias tendências de business intelligence que estão na vanguarda da atualidade, como a capacidade dos sistemas de informação de aprender e prever com base nos interesses do utilizador. Luhn é popularmente conhecido como o pai de Business Intelligence.

Nos anos 70, a mudança começou devido aos avanços tecnológicos e o surgimento de novas formas de se armazenar os dados . Larry Ellison e dois amigos lançaram a primeira versão comercial da base de dados Oracle. Foi o primeiro verdadeiro sistema de gerenciamento de base de dados relacional do mercado. Essa tecnologia ditaria a história e as tendências da BI nas próximas décadas.

Na década de 80, o termo Business Intelligence passou a ser utilizado como uma forma de referir-se ao processo de coleta, organização, análise, compartilhamento e monitoramento de dados.

Nos anos 90, ainda era complicado o cruzamento dos dados, o armazenamento era grande mas para utilizar-se dessa informação era muito complexo. Por volta de 1993 surge a Data Warehouse, um

dos principais componentes do BI. Em 1996 o conceito se alastrou como um processo do Sistema de Informação Executiva, criado pelo MIT no final da década de 70.

No início dos anos 2000, o BI passou a ser uma ferramenta que integrava diversas outras, sendo disponibilizada para um número maior de pessoas. Ele estava em uma preparação para ser o que é hoje. Com a evolução da tecnologia foi-se acrescentando diversas ferramentas, como Data Marts, Data Mining, OLAP, dentre outras.

Com o advento da internet 3.0, tem se tornado difícil garantir que todos os processos de negócio estejam sob controle, uma vez que o enorme volume de dados desafia cada vez mais as empresas. As organizações e empresas precisam de certa ajuda quando se trata de tomar decisões sustentáveis e lucrativas. Ao utilizar ferramentas de BI modernas e profissionais, cada desafio pode ser resolvido, sem a necessidade de um envolvimento massivo de TI.

2.2.1. Estado da Arte

O surgimento de ferramentas de BI revolucionaram o mercado empresarial, que para além de coletar, analisar, monitorar e prever cenários de negócios futuros, cria uma perspectiva clara de todos os dados que uma empresa gerencia.

Atualmente, existem várias ferramentas de implementação e prática de BI. Em Fevereiro de 2020, o Grupo Gartner publicou um relatório sobre as ferramentas de BI líderes no mercado e, por 13 anos consecutivos, o Gartner reconheceu a Microsoft como líder do Magic Quadrant em plataformas de Analytics e BI.

A Microsoft manteve-se sempre no topo devido às ferramentas que oferece tais como o SSIS-SQL server Integration Services para a integração de dados, SSAS - SQL Server Analysis Services para o processamento analítico (OLAP), SSRS-SQL Server Reporting Services para criar, implementar e gerir relatórios e o POWER BI.

A imagem abaixo (figura 8) mostra o posicionamento das empresas com base nas ferramentas de Analytics e BI segundo o estudo da Gartner:



Figura 8-Quadrante Mágico para plataformas de Analytics e BI-(Gartner 2021)

2.2.2. Arquitetura dos Sistemas de BI

Um sistema de BI é um sistema que integra um conjunto de ferramentas e tecnologias que permitem extrair significado de grandes conjuntos de dados. A figura abaixo (figura 4) ilustra a arquitetura de um sistema de BI. Normalmente, um sistema típico de BI é composto pelas seguintes partes: Fontes de dados, ETL, Data warehouse e Análise de dados. Para uma boa implementação e funcionamento, é necessário que a arquitetura do sistema não exclua nenhum desses passos.

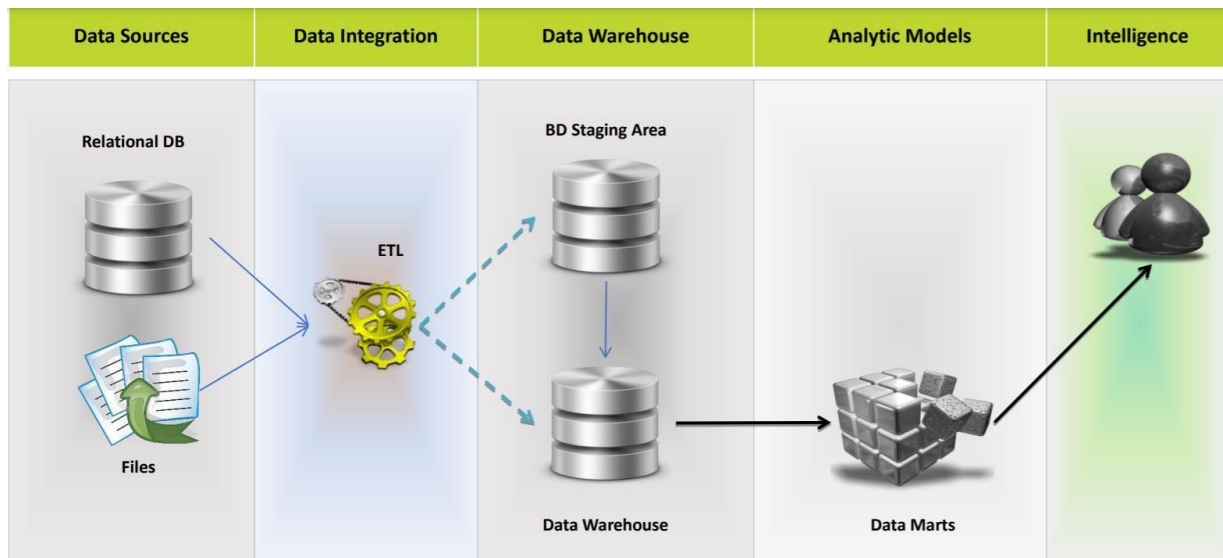


Figura 9-Arquitetura de um Sistema de BI

2.2.2.1. Data Sources :

Um dos grandes desafios em sistemas que baseiam em BI é a fusão de dados. Geralmente esses dados são provenientes de sistemas transacionais (OLTP) e podem estar em vários formatos como xml, xls, txt, csv entre outros.

De acordo com Moss e Atre (2003) destacam-se três principais Fontes de Dados:

- **Interna:** bases de dados interna que contêm informação crítica da empresa;
- **Externa:** dados externos que podem ser adquiridos de vendedores especializados que recolhem informações específicas e disponíveis no domínio público.
- **Privada:** Os dados desta fonte podem ser obtidos em computadores pessoais de colaboradores das organizações.

2.2.2.2. ETL-Extract, Transform e Load

Como já mencionado anteriormente ,ETL significa Extract, Transform e Load. O processo de ETL desempenha um papel fundamental num sistema de BI e está dividido em três etapas como é demonstrado na figura abaixo(figura 10):



Figura 10-Processo ETL

Nas empresas , há sempre um grande volume de dados provenientes de diferentes fontes de dados, mas todos esses dados devem ser integrados no mesmo repositório.

É tentador pensar que a criação de um data warehouse é simplesmente extrair dados de várias fontes e carregá-los numa base de dados de uma data warehouse. Isso está longe de ser verdade e requer um processo ETL complexo. O processo de ETL requer contribuições ativas de várias partes interessadas, incluindo desenvolvedores, analistas, testadores, altos executivos e é tecnicamente desafiador.

A fim de manter seu valor como uma ferramenta para os tomadores de decisão, o sistema de data warehouse precisa mudar com as mudanças nos negócios. ETL é uma atividade recorrente (diária, semanal, mensal) de um sistema de Data warehouse e precisa ser ágil, automatizado e bem documentado.

O processo ETL é dividido em três etapas, como o nome sugere. As três etapas são as seguintes:

Extrair: Nesta etapa da arquitetura ETL, os dados são extraídos do sistema origem para a Staging Area. As transformações, se houver, são feitas nesta Staging Area para que o desempenho do sistema de origem não seja degradado. Esta área de teste oferece a oportunidade de validar os dados extraídos antes que eles sejam movidos para a data warehouse.

A extração dos dados é feita através de fontes internas e externas, também, cada sistema pode usar um formato de dados diferente. Os formatos de fonte de dados comuns são ficheiros simples e base de dados relacionais, no entanto, também podem ser estruturas de base de dados não relacionais, como o Sistema de Gestão de Informação ou estruturas de dados alternativas. A extração desses diferentes tipos de fontes criam uma dificuldade adicional ao processo.

A determinação de quais dados são relevantes é uma tarefa desafiadora, uma vez que os dados contidos no sistema de origem costuma ser muito complexo. Esta escolha é baseada no design da data warehouse que, por sua vez, depende das informações necessárias para o negócio , análise de inteligência e os sistemas de apoio à decisão.

A fim de manter os dados atualizados na data warehouse, os dados devem ser extraídos várias vezes de maneira rotineira. Existem três métodos de extração de dados:

- Extração Completa: copiar todos os dados da origem para a Staging Area ;
- Extração parcial - sem notificação de atualização;

- Extração parcial - com notificação de atualização;

Independentemente do método usado, a extração não deve afetar o desempenho e o tempo de resposta dos sistemas de origem. Esses sistemas de origem são bases de dados de produção ao vivo. Qualquer desaceleração ou bloqueio pode afetar os resultados financeiros da empresa.

Transformar é a próxima etapa no processo ETL. Nesta fase, um conjunto de regras e / ou funções são implementadas para os dados extraídos da fonte em para preparar os dados a serem carregados no destino final. A manipulação de dados depende de sua fonte, isso significa que algumas fontes exigirão poucas transformações enquanto outros exigem bastante.

A melhoria da qualidade dos dados extraídos das diferentes fontes é o principal objetivo da fase de limpeza e transformação, isso é feito como consequência da correção de discrepâncias, valores ausentes e imprecisões.

A última etapa do processo ETL, é o **Load**, carregamento, que consiste em carregar os dados, depois de transformados, na DW. Existem três tipos de carregamento possíveis:

Initial Load: carregamento completo da Data Warehouse.

Incremental Load: carregamento incremental, onde é feito o carregamento dos registros novos e atualizados. Requer menos tempo, mas não há garantia que os novos registros foram atualizados corretamente.

Full Refresh: carregamento completo, onde todas as linhas da tabela são apagadas e os dados são carregados. Este carregamento requer muito tempo mas é confiável, pode-se garantir que o carregamento é realizado com sucesso.

2.2.2.3. Data Warehouse

Após a extração dos dados e as transformações necessárias, esta informação é armazenada em bases de dados designadas por *Data Warehouse* e *Data Mart*. Existem duas abordagens mais utilizadas na construção de um DW que são Inmon (Top-Down) e Kimball (Bottom-Up).

Na metodologia sugerida por Inmon, de cima para baixo, o DW é o elemento central de todo o ambiente analítico. Um data warehouse contém dados transacionais ou altamente detalhados que são extraídos de vários sistemas operacionais e integrados em um modelo de dados empresarial normalizado (Inmon, 2005).

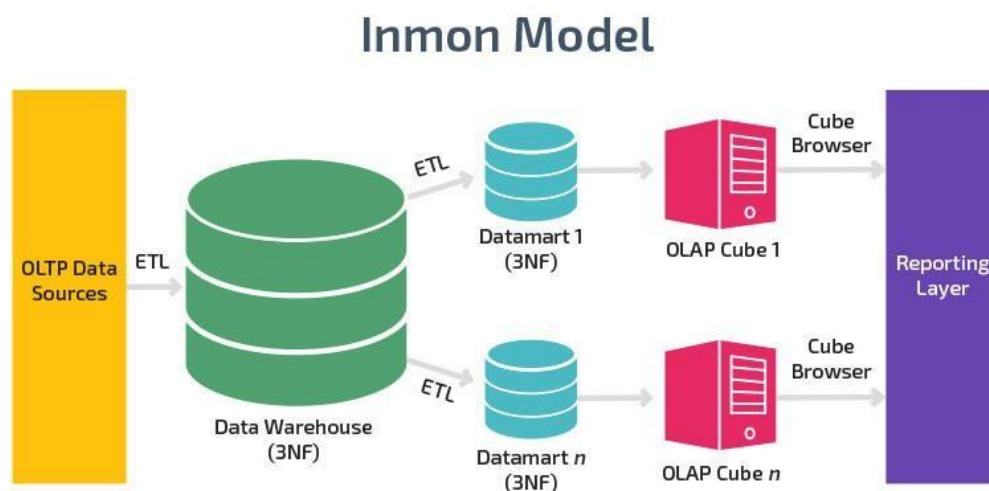


Figura 11-Abordagem de Inmon

De acordo com Inmon (2005), o *Data Warehouse* é um armazém de dados que pode ser caracterizado por:

- **Orientado por Assunto:** os dados são organizados por assuntos específicos e relevantes para a organização. Por exemplo, uma organização pode ter no seu DW informação de clientes, fornecedores, produtos e vendas.
- **Integrado:** o DW geralmente é construída com a integração de dados obtidos de diferentes fontes. Antes da informação ser armazenada no DW, técnicas de integração e limpeza de dados são aplicadas de modo a existir uma representação única no armazém. Ao integrar estes dados no armazém, eles são convertidos de modo a existir uma representação uniforme no *Data Warehouse*.
- **Variante no Tempo:** o *Data Warehouse* mantém os dados históricos (cinco a dez anos) e atuais.
- **Não Volátil:** Uma vez que os dados são armazenados no DW, eles não podem ser alterados. Portanto, dados históricos num *Data Warehouse* nunca devem ser alterados.

A metodologia de Kimball, de baixo para cima, consiste na criação de Data Marts, para formar o DW da organização através de um interface de ligação *bus*.

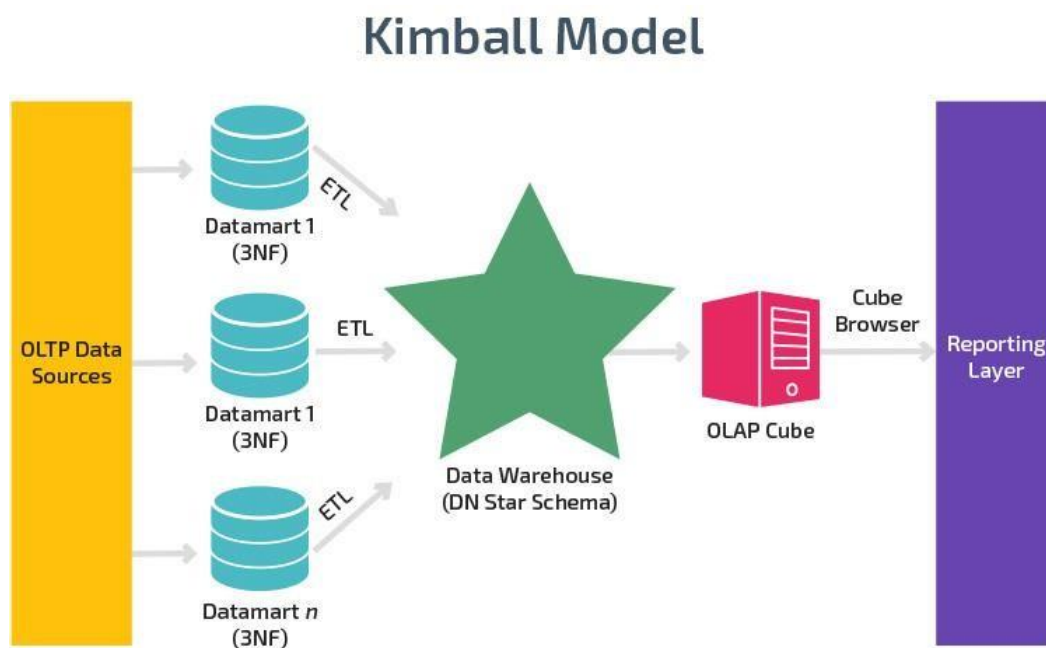


Figura 12-Abordagem de Kimball

2.2.2.4. Analytic Models : Data Mart

Um Data Mart é uma estrutura de dados multidimensional orientada por assunto ou por departamento tais como o departamento de marketing, departamento financeiro, entre outros (Kimbal, 1998). Este processo de agregar a informação em assuntos possibilitará que todos os utilizadores da área em questão tenham acesso à informação de uma forma mais ágil e rápida (Malinowski e Zimányi, 2009).

Os Data Marts podem seguir um dos seguintes esquema :

Star Schema ,esquema em estrela, desenvolvido por Kimball, consiste em tabelas de dimensão e tabela de Fato. As tabelas de dimensão fornecem um contexto em torno de um evento de processo de negócios. Em termos simples, elas dão quem, o quê , quando e onde de um evento. Fatos são as medidas e métricas do processo de negócios. A tabela de fatos contém as medidas numéricas produzidas por um evento operacional . Além de medidas numéricas, uma tabela de fatos sempre contém chaves estrangeiras para cada uma das suas dimensões associadas.

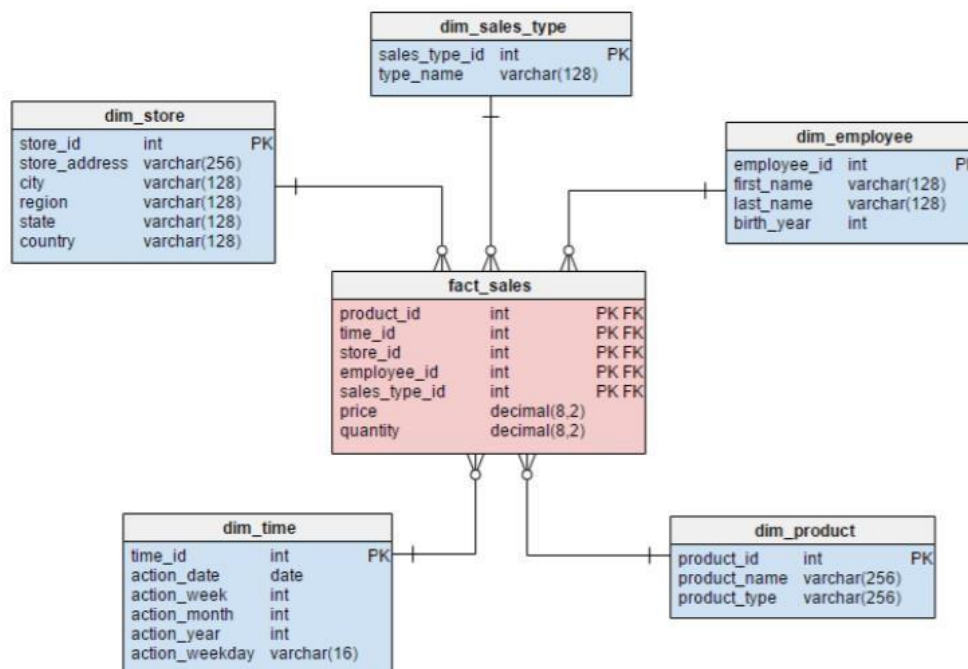


Figura 13-Exemplo Star Schema

Snowflake Schema ,esquema Floco de Neve, é semelhante ao esquema em estrela, no entanto, consiste em tabelas de fatos centralizadas que estão conectados a várias dimensões. Isso significa que as dimensões são normalizadas em várias tabelas relacionadas em contraste com as dimensões desnormalizadas no esquema em estrela, onde cada, dimensão é representada por uma única tabela.

O esquema do floco de neve foi nomeado assim desde a disposição das tabelas no painel multidimensional especificamente o diagrama de relacional das entidades parecer a forma de um floco de neve. Isso é devido ao fato de que as dimensões podem ser compostas por hierarquias de atributos, que não são normalizadas, suas a normalização resulta nesta forma específica.

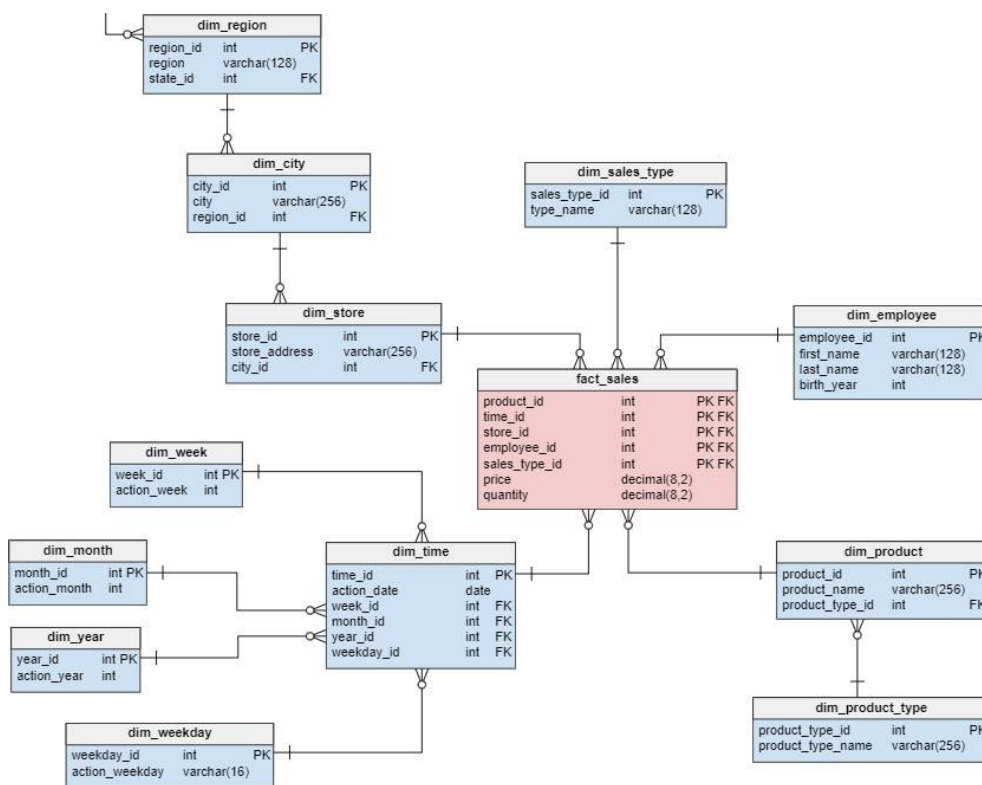


Figura 14-Exemplo Snowflake schema

Os modelos multidimensionais oferecem suporte ao Business Analytics, fornecendo aos analistas da organização dados otimizados para a manipulação e análise. Os modelos multidimensionais são estruturados em torno de métricas, dimensões, hierarquias, e cubos em vez de tabelas, linhas, colunas e relações.

O Online Analytical Processing (OLAP) é uma categoria de software que permite aos usuários analisar informações de vários sistemas de base de dados ao mesmo tempo. É uma tecnologia que permite aos analistas extraírem e visualizarem dados de negócios de diferentes pontos de vista.

Os analistas frequentemente precisam agrupar, agregar e juntar dados. Com o OLAP, os dados podem ser pré-calculados e pré-agregados, tornando a análise mais rápida.

As bases de dados OLAP são divididas em um ou mais cubos. Os cubos são projetados de forma que a criação e a visualização de relatórios se tornem fáceis.

No centro do conceito OLAP, está um Cubo OLAP. O cubo OLAP é uma estrutura de dados otimizada para uma análise de dados muito rápida .

O Cubo OLAP consiste em fatos numéricos chamados medidas que são categorizados por dimensões. Uma dimensão pode ser qualquer visão do negócio que faça sentido para a análise, como produto, departamento. Este modelo de dados multidimensional simplifica para os usuários o processo de formular pesquisas complexas, criar relatórios, efetuar análises comparativas, e visualizar subconjuntos (*slice*) de maior interesse. Por exemplo, um cubo contendo informações de vendas poderá ser composto pelas dimensões tempo, região, produto, cliente e medidas. Medidas típicas seriam valor de venda, unidades vendidas, custos, margem, etc.

Dentro de cada dimensão de um modelo OLAP, os dados podem ser organizados em uma hierarquia que define diferentes níveis de detalhe. Por exemplo, dentro da dimensão tempo, podemos ter uma hierarquia representando os níveis anos, meses, e dias. Da mesma forma, a dimensão região poderá ter os níveis país, região, estado e cidade. Assim, um usuário visualizando dados em um modelo

OLAP irá navegar para cima (*drill up*) ou para baixo (*drill down*) entre níveis para visualizar informação com maior ou menor nível de detalhe sem a menor dificuldade.

As possíveis operações que podem ser realizadas em um modelo OLAP são:

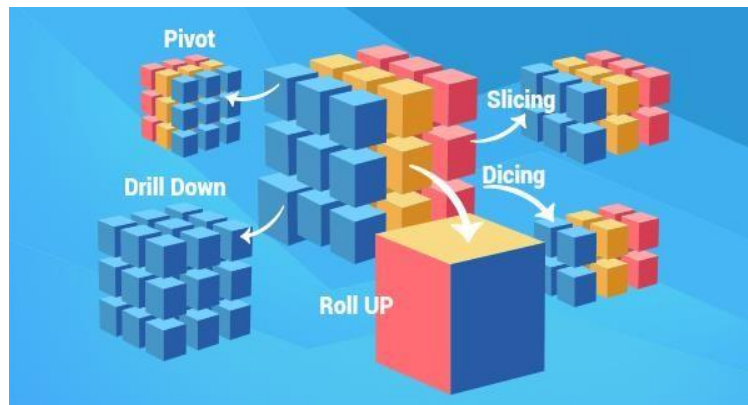


Figura 15-Operações em OLAP

SLICE: selecionar dados de uma única dimensão;

DICE: extrai um sub cubo da seleção de duas ou mais dimensões;

DRILL DOWN: examina dados com maior detalhe;

ROLL UP: combinação de células de uma ou mais dimensões para atingir um nível maior de generalização;

PIVOT: visualiza dados por uma nova perspectiva.

2.2.2.5. Intelligence:Reporting

Os relatórios são uma parte fundamental do BI . que se concentra na visualização de dados em diferentes tipos de visualizações, como tabelas, gráficos e tabelas. As visualizações no contexto de relatórios são uma representação gráfica dos dados, cujo objetivo é apresentar as informações com precisão em um formato que seja digerível para os utilizadores finais.

No passado, as visualizações criadas para fins de relatório eram estáticas, o que significa que os dados não podiam ser manipulados na própria visualização. Como as necessidades dos negócios mudaram com o tempo, os relatórios também evoluíram. No mundo atual dos negócios em rápida mudança, cada vez mais interatividade se tornou essencial para os executivos que precisam tomar decisões de negócios rápidas e dinâmicas.

Nos últimos anos, a interatividade do utilizador final deu um passo à frente com os relatórios ad hoc. Os relatórios ad hoc permitem que os utilizadores finais criem relatórios do zero e também editem e manipulem relatórios predefinidos. A capacidade de criar, editar e guardar relatórios para uso futuro capacitou os usuários mais do que nunca, permitindo a visualização mais rápida de dados pertinentes.

3. TECNOLOGIAS E FERRAMENTAS

3.1. TECNOLOGIAS :

3.1.1. SQL- Structured Query Language:

É uma linguagem de programação que tem como objetivo principal o processamento de dados. De acordo com Damas (2005) , o objetivo da linguagem SQL é servir de interface entre o utilizador dos dados e um SGBDR como o Oracle ou o SQL Server.

3.1.2. PL-SQL - Procedural Language/Structured Query Language

É uma extensão da linguagem padrão SQL , que permite que a manipulação de dados seja incluída em unidades de programas. Blocos de PL/SQL são passados e processados por uma PL/SQL Engine que pode estar dentro de uma ferramenta Oracle ou do Server.

3.2. FERRAMENTAS:

3.2.1. Microsoft SQL SERVER :

O SQL Server é um sistema gerenciador de Base de dados relacional (SGBD) desenvolvido pela Microsoft e que tem como objetivo armazenar dados e recuperar dados solicitado por outros aplicações.

Ele fornece suporte para operações de BI e analítica e também para uma diversidade de processos de transação. Como o nome sugere, a linguagem usada nesta ferramenta é SQL.

O SQL Server tem quatro elementos principais:

- Database Engine : este elemento serve para armazenar dados OLTP.
- Reporting Services: SSRS é uma componente de Business Intelligence que permite criar relatórios através dos dados transformados e lidar com o acesso concedido aos relatórios.
- Serviços de Integração: SSIS gerencia o processo ETL, o que significa que auxilia no processo de transformação de dados em um formato padrão para ser usado por todos os elementos da organização.
- Analysis Services: SSAS incorpora uma ferramenta OLAP para analisar os dados online, bem como um motor de mineração de dados que ajuda a encontrar os dados solicitado pelas consultas.

3.2.2. Visual Studio Team system:

Visual Studio Team System Team Foundation Server (TFS) ajuda às equipas de desenvolvimento de software pois , as equipas ganham autonomia para realizar um trabalho com serviços de controle de versão, construção e relatórios integrados em uma plataforma de colaboração única e segura.

3.2.3. Pentaho Data Integration:

O Pentaho é formado por um conjunto de softwares que servem para a construção de soluções de BI, que incluem programas para extrair os dados de sistemas de origem e gravá-los numa data warehouse (ou base de dados), limpá-los, prepará-los e entregá-los a outros sistemas de destino . O Pentaho Data Integration (PDI) ,também conhecido como Kettle (Kettle Extraction Transformation Transport Load Environment) é uma parte do Pentaho Open Source Business intelligence suite. Inclui software para todos os aspetos de apoio à tomada de decisões de negócios.

Pentaho Data Integration é conhecido pela sua facilidade de uso e rápida aprendizagem .

O Pentaho permite que administradores e desenvolvedores de ETL criem seus próprios trabalhos de manipulação de dados com um criador gráfico amigável (Spoon) e sem inserir uma única linha de código.

O PDI usa um repositório comum e compartilhado que permite a execução remota do ETL , facilita o trabalho de equipa e simplifica o processo de desenvolvimento.

Com o Pentaho Data Integration é possível fazer inúmeras operações de Integração de Dados. Como por exemplo:

- Migração de dados;
- Movimentação de grandes volumes de dados;
- Transformação de dados;
- Limpeza de dados;
- Conformidade de dados, entre outros.

O Pentaho Data Integration é formado por duas categorias de artefactos, Jobs e Transformações, e estes artefactos são construídos por meio da sua interface gráfica, o Spoon. O Spoon é a interface gráfica do Pentaho Data Integration que facilita na conceção de rotinas e lógica ETL.

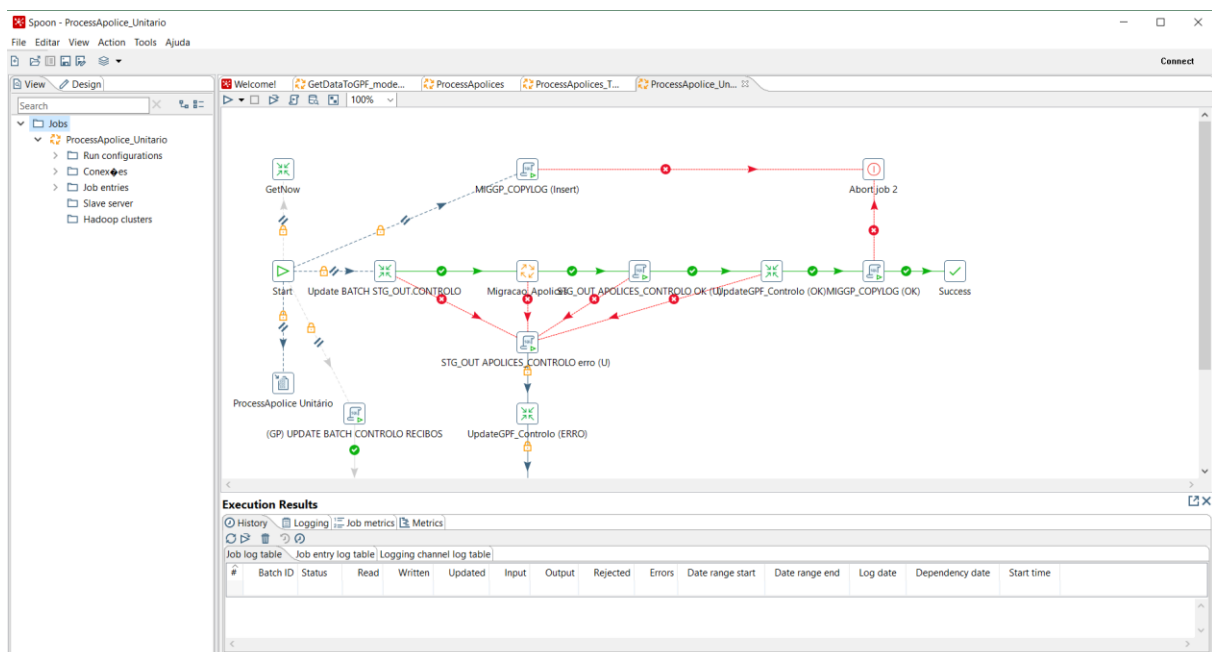


Figura 16-Interface do Spoon

Uma transformação regista o passo-a-passo de como a extração ou a leitura de uma fonte de informação é realizada. É a transformação que opera sobre os dados. Ela pode conter:

- Leitura de dados de uma tabela de uma base de dados;
- Seleção de campos específicos de uma tabela;
- Concatenação de valores de dois campos distintos de uma tabela;
- Fórmulas matemáticas ;
- Merge de dados de tabelas contidas em base de dados diferentes;
- Merge de dados originados em tabelas, ficheiros XML, TXT ou CSV, entre outras fontes de dados;
- Aplicação de expressões regulares em texto para limpeza e filtros.

O aspecto mais importante numa transformação é que ela opera todas as etapas simultaneamente, uma transformação não tem início ou fim, ela apenas processa linhas que chegam.

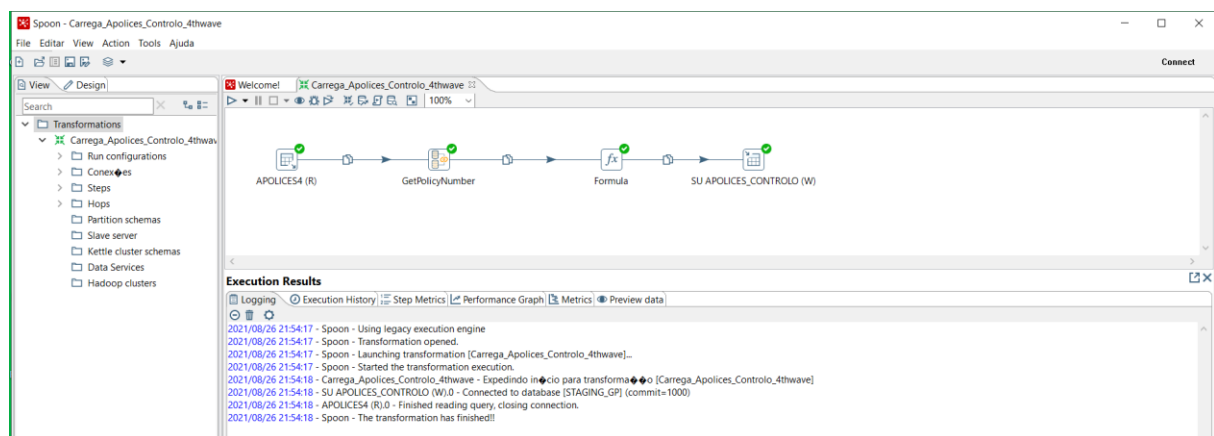


Figura 17-Exemplo de uma transformação no Spoon

O Spoon, é apenas o interface gráfico para criar os processos de integração de dados. Ele não serve para executá-los em produção. O Pan executa as transformações modeladas em Spoon.

Enquanto o Pan executa transformações, o Kitchen executa jobs.

Um Job sequencia operações. Ao contrário de uma transformação, um job realiza operações completas, uma por uma. Ele permite, por exemplo, combinar transformações numa sequência específica e, com isto, automatizar uma dada tarefa. Por sua natureza, ele não fornece muitos recursos técnicos para manusear os dados em si, deixando isto à cargo das transformações.

A tríade Spoon, Pan e Kitchen são os responsáveis pela criação e execução de artefactos criados para solucionar um problema de extração, transformação e carregamento de dados num projeto de ETL com o Pentaho Data Integration.

4. METODOLOGIA

Sem dúvida, uma metodologia devidamente definida e estritamente seguida para a gestão de um projeto fornece uma garantia firme que o trabalho será feito no prazo, dentro do orçamento e de acordo com as especificações do cliente.

O projeto de Integração da Generali Portugal (GP) nas Seguradoras Unidas(SU) assenta em três pilares muito importantes:

- RH e Organização – com um claro foco nas pessoas e na criação de uma estrutura organizacional única e preparada para os novos desafios;
- Proposta de Valor – com o objetivo de oferecermos uma proposta de valor única aos nossos clientes e parceiros de negócio, tendo em vista um elevado grau de qualidade e inovação;
- Sistemas e Processos – com o compromisso de dispormos das melhores ferramentas tecnológicas, soluções e processos para agilizarmos a integração e a nossa transformação.

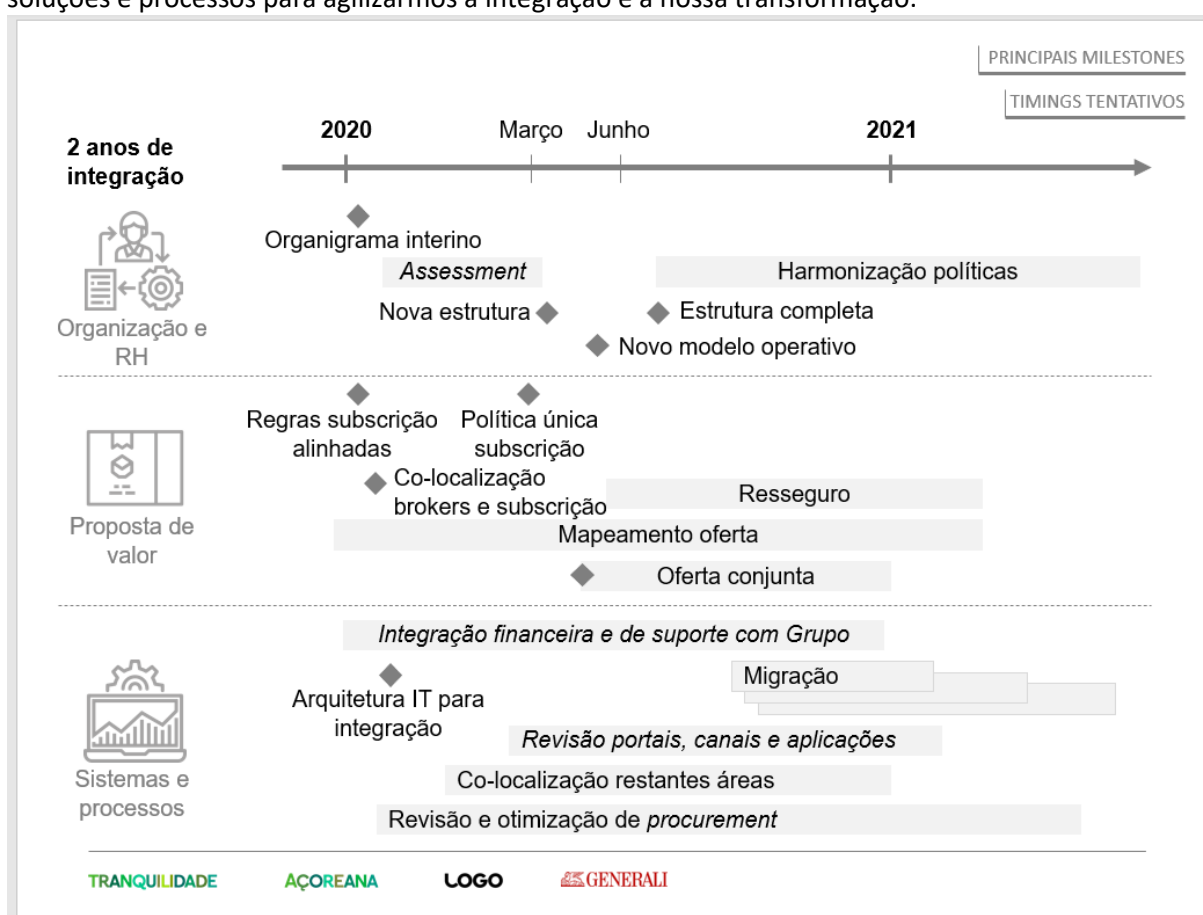


Figura 18-Objetivos do Projeto Integração

Para agilizar a sua implementação, o projeto de integração foi dividido em 5 fases:

1. Lançamento (Janeiro);
2. Desenho (Fevereiro a Abril);
3. Abordagem única (Maio a Setembro);
4. Sistemas únicos (Outubro em diante);
5. Descomissionamento.

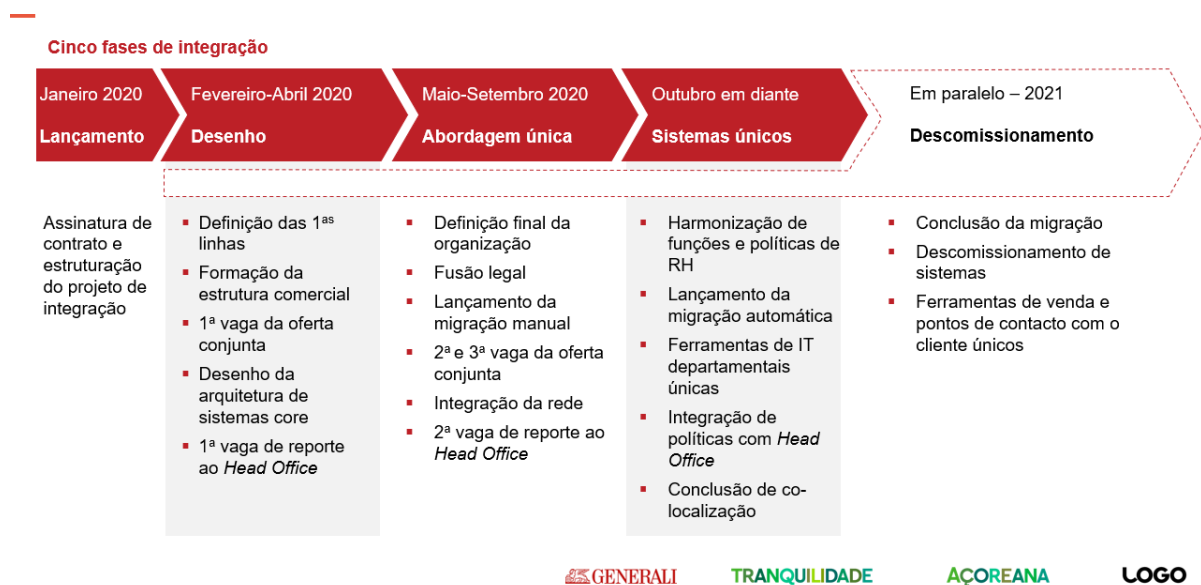


Figura 19-As cinco fases de Integração

Para manter um ritmo constante e alcançar os milestones definidos de forma consistente, o modelo de governance do Projeto de Integração é composto por três ciclos distintos:

- Reuniões mensais entre os líderes da Generali Internacional, Comité Executivo e PMO (Project Management Office) para seguimento das atividades, status dos workstreams prioritários e captura de sinergias, e decisões-chave;
- Reuniões semanais entre Comité Executivo, PMO e co-heads dos workstreams para seguimento do plano de atividades, status das workstreams prioritárias, e tomada de decisão;
- Interações regulares semanais e diárias, geridas pelos co-heads de cada workstream, e com PMO, garantindo o progresso das ações.

Este projeto foi inicialmente (Janeiro 2020) estruturado em 13 workstreams (grupos de trabalho), organizados por áreas de atuação, com equipas conjuntas da GP e da SU, para ser alcançado mais de 500 atividades e 336 milestones.

Os 13 grupos de trabalho que compõem o Projeto de Integração estão orientados para os seguintes temas: Comunicação; RH e organização; TI; Migração manual; Oferta, subscrição e resseguro; Comercial; Operações e sinistros; *Procurement* e instalações; Fusão legal; Finanças, fiscal e atuariado; Investimentos; Jurídico, *compliance* e outras funções de suporte e Sinergias.

Nas figuras abaixo são apresentados o âmbito de cada workstream, bem como os seus co-líderes locais e coordenadores do *Project Management Office (PMO)*:

Workstream / Grupo de Trabalho	Âmbito	Co-líderes locais		Integração PMO – equipa central
		SU	GP	
1 Comunicação	<ul style="list-style-type: none"> Comunicação interna e externa (incluindo parceiros de distribuição) Abordagem de imagem e comunicação 	<ul style="list-style-type: none"> André Taxa 	<ul style="list-style-type: none"> Conceição Tomás 	<ul style="list-style-type: none"> Alexandra Vaz
2 RH e organização	<ul style="list-style-type: none"> Organização única e co-localização das equipas 	<ul style="list-style-type: none"> Gilda Raposo 		<ul style="list-style-type: none"> João Gonçalves
3 TI	<ul style="list-style-type: none"> Única arquitetura e estrutura de TI Migração de sistemas 	<ul style="list-style-type: none"> Pedro Santos 	<ul style="list-style-type: none"> Juan Mederos 	<ul style="list-style-type: none"> Ana Pereira Alexandra Vaz Pedro D. Ferreira
4 Migração manual	<ul style="list-style-type: none"> Migração das políticas inviáveis para automação 	<ul style="list-style-type: none"> João Marques 	<ul style="list-style-type: none"> Ana Pinto 	
5 Oferta, subscrição e resseguro	<ul style="list-style-type: none"> Oferta e regras de subscrição únicas Mapeamento de oferta para o processo de migração Alinhamento das ferramentas de venda 	<ul style="list-style-type: none"> Carlos Silva Luís Ribeiro Isabel Farinha João Lopes Pedro Rosário Luís Rodrigues Artur Mirones Luís Prazeres Vera Barros 	<ul style="list-style-type: none"> Joaquim Aguiar Paulo Mascarenhas Giampiero Prester Rogério Santos Giacomo Bertoldi 	<ul style="list-style-type: none"> Ana Pereira Marta Colaço
6 Comercial	<ul style="list-style-type: none"> Integração de corretores, agentes e outras estruturas comerciais Modelo único de remuneração Modelo único de cobertura Ferramentas e processos únicos 	<ul style="list-style-type: none"> João Amaral Emmanuel Lesueur Carlos Cotrim Tiago Gomes 	<ul style="list-style-type: none"> Alberto Severino Orlando Cardoso 	<ul style="list-style-type: none"> Miguel Brandão

Figura 20- Workstreams Janeiro 2020

Workstream / Grupo de Trabalho	Âmbito	Co-líderes locais		Integração PMO – equipa central
		SU	GP	
7 Operações e sinistros	<ul style="list-style-type: none"> Modelo de operação e definição da capacidade Harmonização de processos e ferramentas de suporte 	<ul style="list-style-type: none"> João Marques 	<ul style="list-style-type: none"> Ana Pinto 	<ul style="list-style-type: none"> Cláudia Silva
8 Procurement e instalações	<ul style="list-style-type: none"> Alinhamento com procedimentos GHO Otimização dos esforços de procurement Co-localização das equipas e Gestão dos bens imobiliários 	<ul style="list-style-type: none"> Miguel Pereira Paulo Ferreira 	<ul style="list-style-type: none"> Assur Della Torre 	<ul style="list-style-type: none"> Cláudia Silva
9 Fusão legal	<ul style="list-style-type: none"> Fusão legal das três entidades 		<ul style="list-style-type: none"> Workstream gerido pela Região 	
10 Finanças, fiscal e atuariado	<ul style="list-style-type: none"> Integração com ferramentas e procedimentos GHO Processos locais e harmonização de reportes (incl. conceitos e fontes de informação) 	<ul style="list-style-type: none"> Pedro Medalhas Maria Duarte Tiago Coelho Miguel Pereira Ana Ferreira Luís Rego 	<ul style="list-style-type: none"> António Silva Marta Nabais 	<ul style="list-style-type: none"> Ana Ferreira
11 Investimentos	<ul style="list-style-type: none"> Alinhamento com ferramentas e processos GHO Configuração do novo modelo operacional 	<ul style="list-style-type: none"> Francisco Marques 	<ul style="list-style-type: none"> Nuno Megre 	<ul style="list-style-type: none"> Ana Ferreira
12 Jurídico, compliance e outras funções de suporte	<ul style="list-style-type: none"> Alinhamento com ferramentas e processos GHO Coordenação da estrutura local 	<ul style="list-style-type: none"> Manuela Simões Natasha Revez Paulo Costa Rui Miranda Francisco Dias 	<ul style="list-style-type: none"> José Passos João Terlica Jorge Lima Filipa Moreira 	<ul style="list-style-type: none"> Ana Ferreira
13 Sinergias	<ul style="list-style-type: none"> Definição de targets e Gestão do modelo financeiro "KPIs físicos" e planos operacionais Monitorização da captura de valor 	<ul style="list-style-type: none"> Miguel Pereira/Ana Ferreira (modelo financeiro) João Gonçalves (plano operacional) 	<ul style="list-style-type: none"> António Silva 	<ul style="list-style-type: none"> João Gonçalves

Figura 21-Workstreams Janeiro 2020 (continuação)

Em Junho de 2020 , foi desenhada uma nova estrutura organizativa da empresa tendo em conta os objetivos e estratégia do Grupo Generali, mas também tendo em consideração às necessidades e especificidades, bem como os pontos fortes, experiência e competência da equipa.

Após a reestruturação organizacional , o projeto foi dividido em 11 grupos de trabalho com as seguintes responsabilidades :

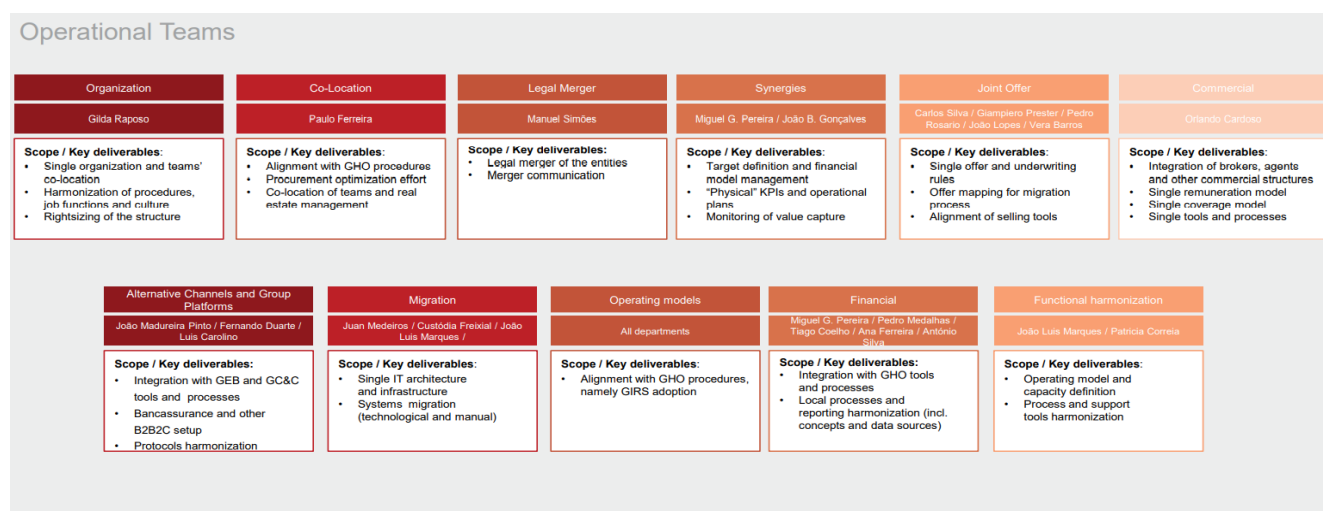


Figura 22-Workstreams Junho 2020

Posteriormente foi definido o Modus Operandi da Migração Tecnológica . A figura abaixo(figura 23) representa o plano de dois anos da integração .

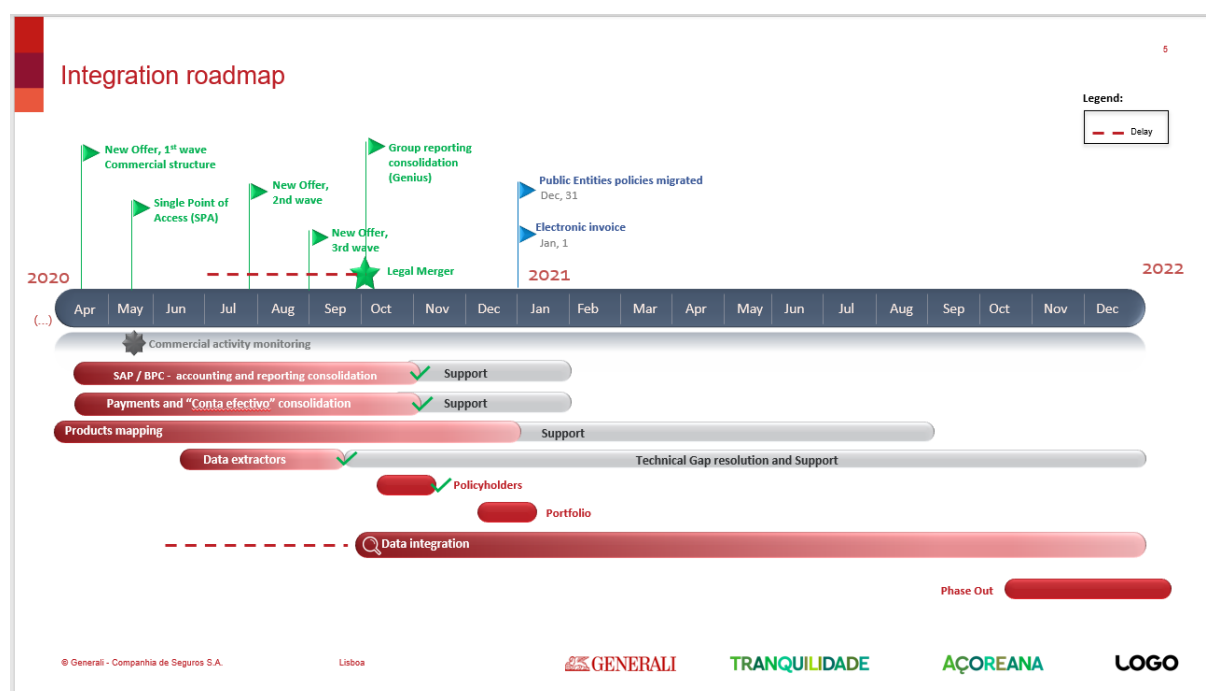


Figura 23-Macro Cronograma do Projeto de Integração

Durante o Mapeamento dos Produtos ,foi feito um estudo preliminar dos tipos de produto existentes da Generali . Os seguros dividem-se em duas categorias distintas: Ramo Vida e Ramo não vida.

Os seguros do ramo vida engloba os seguros de vida clássicos e os seguros financeiros: seguros de capitalização e PPR (Planos Poupança Reforma). Estes produtos serão migrados para o sistema AS400 da Tranquilidade .

No ramo não vida englobam-se todos os seguros que têm objeto bens patrimoniais tais como: Acidentes Pessoais, Acidentes de trabalho, Responsabilidade civil automóvel, Seguro de Automóvel, Multirriscos Habitação, Multirriscos Condomínio, Seguro de caça entre outros. Os novos seguros que vão sendo criados pertencem normalmente ao ramo não vida, pois visam proteger bens patrimoniais. Estes produtos do Ramo Não vida serão migrados para o Sistema destino Cogen da Tranquilidade. O quadro abaixo mostra a quantidade total de apólices existentes por tipo de produtos:

Grupo Produtos	Migração Manual			Migração tecnológica			Total # apólices
	# Apólices	Peso (% total)	Cód. Produtos	# Apólices	Peso (% total)	Cód. Produtos	
Automovel	2 105	1%		171 703	52%		173 808
Individual	470	0%	81, 82, 83	171 703	52%		172 173
Frotas	1 635	0%	8084, 9084	0	0%		1 635
Incendio e RD	4 825	1%		66 249	20%		71 074
Multirriscos	933	0%	312, 313,318,319, 516, 9300	60 490	18%		61 423
Engenharia	878	0%	92, 140,141,142,143,144,145,146	5 759	2%		6 637
Incendio e L.C.	2 833	1%	90, 91, 9090	0	0%		2 833
Outros patrimoniais	181	0%	86, 100, 101,102, 103, 180, 181, 9200	0	0%		181
Acidentes Trabalho	0	0%		28 566	9%		28 566
Acidentes de Trabalho	0	0%		18 937	6%		18 937
Trabal. Independentes	0	0%		9 629	3%		9 629
Acidentes Pessoais	1 083	0%	2, 3, 4, 5, 6, 9, 30, 8004	13 107	4%		14 190
Saude	9 500	3%		0	0%		9 500
Individual	8 366	3%	21, 22, 23, 24	0	0%		8 366
Grupo	1 134	0%	9020, 9021, 9022,9024	0	0%		1 134
Resp. Civil	164	0%	152, 154, 158	9 735	3%		9 899
Transportes	566	0%	71, 72, 74, 75, 76, 77	0	0%		566
Risco	424	0%		7 695	2%		8 119
Individual	0	0%		7 695	2%		7 695
Grupo	424	0%	8120, 8520, 9120, 9220, 9520	0	0%		424
Universal life	150	0%	29, 229, 9029	0	0%		150
Total	18 817	6%		297 055	89%		315 872
Capitalizacão	0	0%		17 282	5%		17 282
Individual	0	0%		17 157	5%		17 157
Grupo	0	0%		125	0%		125
Total	0	0%		17 282	5%		17 282
Total	18 817	6%		314 337	94%		333 154

Tabela 1-Análise Preliminar das Apólices

O processo de integração de dados está dividido em duas migrações : a manual e a tecnológica (automática). A equipa em que estive integrada apenas era responsável pela migração automática, portanto os processos da migração manual foram elaborados por uma outra equipa da empresa.

Portfolio integration strategy

Data migration high-level life cycle

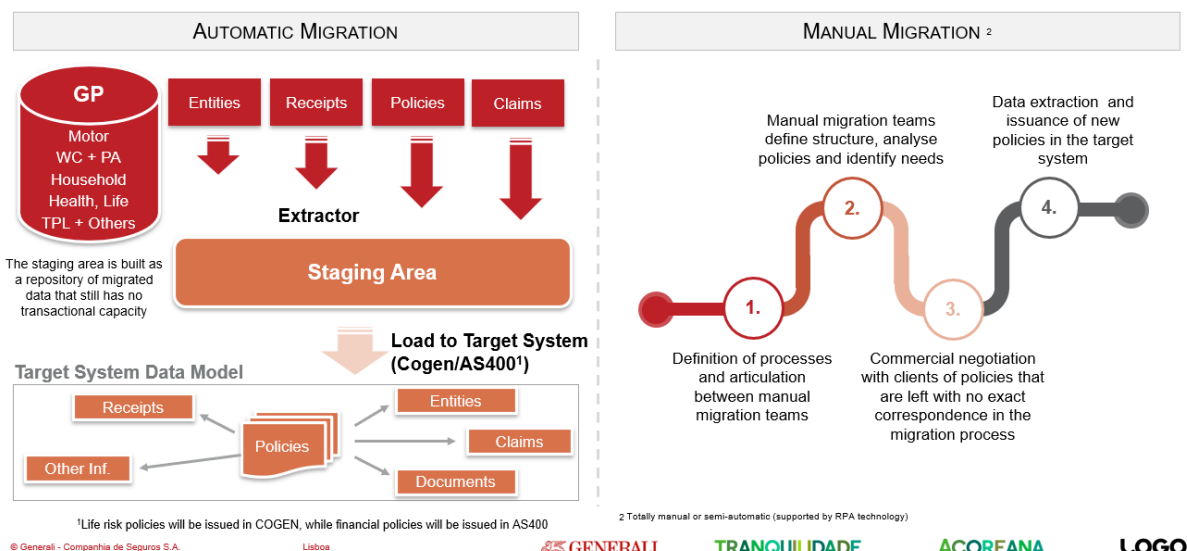


Figura 24-Migração Automática e Manual

A migração Automática consiste na extração dos dados das Entidades, Recibos, Apólices e Sinistros da base de Dados da GP para uma Staging area temporária. A Staging Area é um local temporário onde os dados dos sistemas de origem são copiados.

Esta base de dados é necessária principalmente numa Arquitetura de Data Warehouse por motivos de tempo. Todos os dados necessários devem estar disponíveis antes que os dados possam ser integrados à Data Warehouse.

Devido aos diversos ciclos de negócios, ciclos de processamento de dados, limitações de recursos de hardware e rede e fatores geográficos, não é viável extrair todos os dados de todas as bases de dados operacionais exatamente ao mesmo tempo. Portanto, em razão disso, a extração dos dados da migração automática foi dividida em tranches:

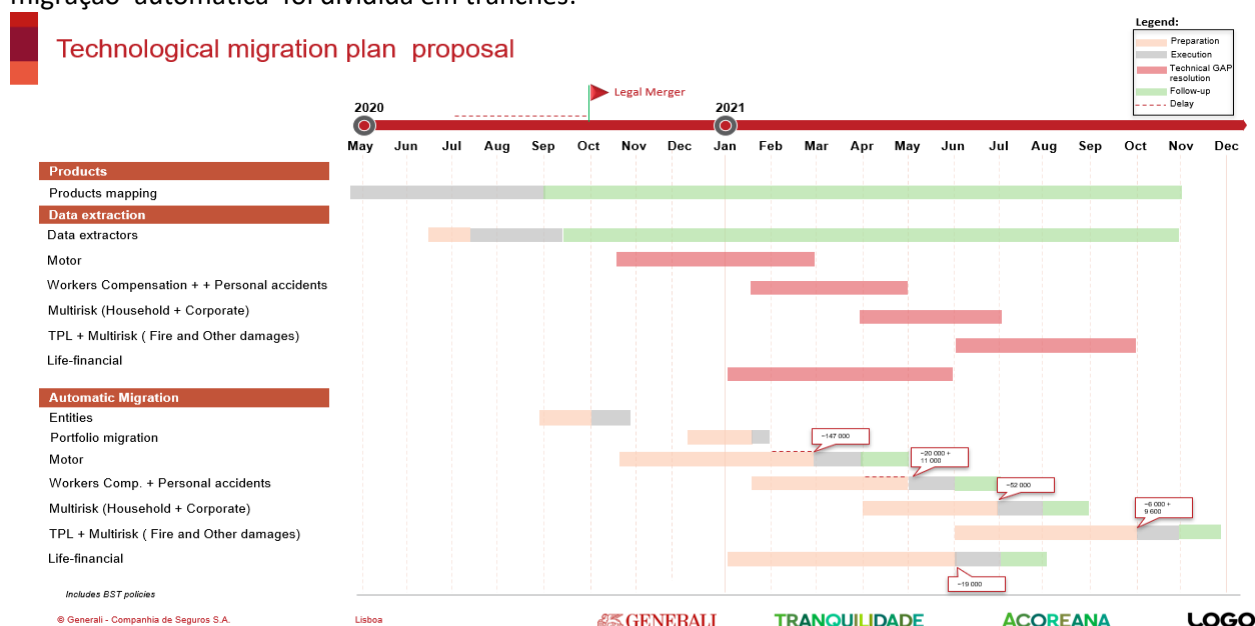


Figura 25-Cronograma da Migração Automática

Quando iniciei o estágio nas últimas semanas de Setembro , as Entidades (Clientes/Tomadores) da Generali já tinham sido migradas no ambiente de produção . Portanto no início do mês de Outubro iniciámos o desenvolvimento dos ETLs para a migração de 3 Produtos de Automóvel da primeira tranche ,cerca de 453 mil apólices :

1. Produto 0080 Automóveis
2. Produto 0084 - Generali Automóvel
3. Produto 8084 - Generali Automóvel Grupo Aberto

Estes produtos foram migrados no início do mês de Março e em seguida demos início à preparação da migração da segunda tranche correspondente aos Produtos:

1. Produto 0001 Acidentes Pessoais
2. Produto 0002 Acidentes Pessoais Escolar
3. Produto 0003 Acid. Pessoais Viagem
4. Produto 0004 Acid. Pessoais Viagem
5. Produto 0005 Acidentes Pessoais Desporto, Cultura e Recreio
6. Produto 0006 Acidentes Pessoais Familiar
7. Produto 0007 Acidentes Pessoais Valor Fixo
8. Produto 0008 Acidentes Pessoais Congressista
9. Produto 0009 Acidentes Pessoais
10. Produto 0010 Generali Agência Viagens
11. Produto 0011 Generali Universidade
12. Produto 0012 Acidentes de Trabalho
13. Produto 0014 Seguro de área (AT)
14. Produto 0015 Avenças Agrícolas (AT)
15. Produto 0016 Trabalhadores Independentes
16. Produto 0030 Acidentes Pessoais Proteção Familiar
17. Produto 8004 Acidentes Pessoais Viagem - Grupo

O estágio terminou logo após a conclusão da Migração da segunda tranche (Maio 2020).

5. TRABALHO DESENVOLVIDO

O trabalho de estágio consistiu em criar de raiz a Staging area STAGING_GP tendo por base a uma base de dados intermediária da Generali a Gensudw. Primeiramente , foi necessário identificar as principais entidades e relações existentes no Koolsite. A sequência de migração de dados é, respetivamente: entidades/clientes, a apólice, recibos (todos os existentes numa apólice dentro de 10 anos ou em existência) e por fim os riscos e sinistros .Esta ordem permite criar as relações entre os vários objetos relacionados com o valor da apólice e os recibos emitidos permitindo migrar uma apólice em estrela.

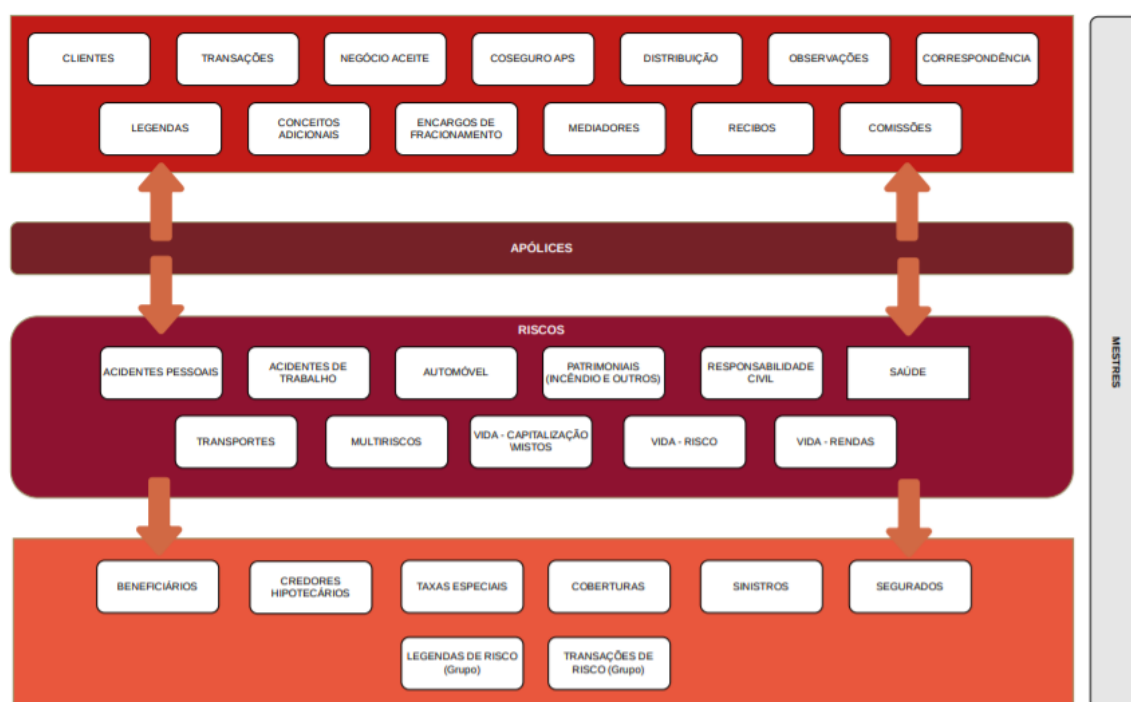


Figura 26-Modelo de Dados da Generali Portugal

Em seguida foi feito o levantamento das tabelas a serem criadas na BD :

Os dados das Entidades/Clientes da Generali estão apenas numa tabela a SPA_CLIENTES , com uma chave primária ,o nº Cliente , e que contém a seguinte informação:

Clientes							
Dados Básicos		Contactos		Moradas - lista		Fusionamentos - lista	
Particular	Colectivo	Particular	Coletivo	Particular	Colectivo	Particular	Colectivo
Tipo	Tipo	Nº Telef Residência	Nº Telef Principal	Tipo Morada	Tipo Morada	Nº Cliente	Nº Cliente
Situação	Situação	Nº Telemóvel	Nº Telef Secundário	Morada	Morada		
Nº Cliente	Nº Cliente	Nº Telef Emprego	Nº Fax Principal				
Nome	Nome	Nº Fax Residência	Nº Fax Secundário				
Data Nascimento	Sigla	Nº Fax Emprego	Email				
Sexo	NIF	Email	Contacto Preferencial				
NIF	CAE	Contacto Preferencial	Período Preferencial				
Sufixo		Período Preferencial	Hora Contacto				
Título Profissional		Hora Contacto					
Nacionalidade							

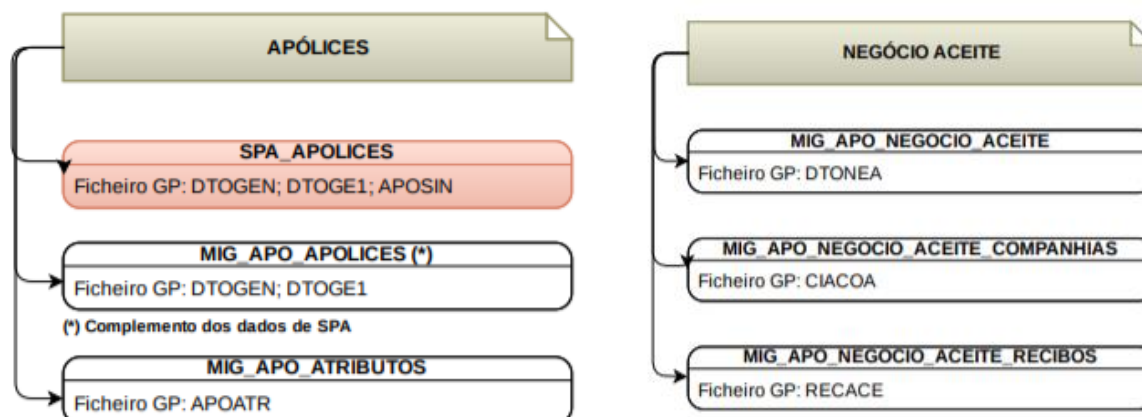
Tabela 2-Informação do Cliente

A informação relevante das apólices encontra-se na tabela SPA_APOLICES em que as chaves primárias dessa tabela são o CODFAM que corresponde ao código da companhia :1 para produtos do Ramo Vida e 2 para os produtos do Ramo Não Vida. Para além do CODFAM a outra chave desta tabela é o NUMINT (número interno da apólice).

Apólices									
Dados Básicos		Relações - lista		Objecto(s) Seguro(s) - lista					
Não Vida	Vida	Particular	Coletivo	Auto	Quadro Pessoal	Local de Risco	Item	Geral	
Nº Apólice	Nº Apólice	Tipo/Descrição	Tipo/Descrição	Matrícula	Tipo Quadro	Morada	Item	Quantidade	
Situação	Situação	Nº Cliente	Nº Cliente	Data Matrícula	CAE	Código Postal	Valor	Unidade Risco	
Produto	Início	NIF	NIF	Marca		Freguesia	Descrição	Capital	
Opção	Vencimento	Nome	Nome	Modelo		Concelho			
Início	Modo Renovação	Data Nascimento	Sigla	Versão		Distrito			
Vencimento	Período	Sexo	CAE	Cilindrada		País			
Modo Renovação	Início do Período			Potência		Ano Construção			
Duração	Fim Período			Lugares		Capital			
Período	Prémio Período			Peso Bruto		Área Bruta (m2)			
Situação Período	Fraccionamento			Data Construção		Nº Assoalhadas			
Início do Período				Combustível					
Fim Período									
Início Transacção									
Fim Transacção									
Situação Transacção									
Tipo Transacção									
Prémio Comercial									
Prémio Total									
Método Pagamento									
Fraccionamento									
Data de Renovação									
Conta distribuição									
Conta Cobrança									
Conta Captação									
Conta Prestação de Serviços									
Conta Técnico Comercial									
Conta adicional									
Data Cancelamento									
Motivo Alteração									
Morada para Documentação									
Descrição Objeto Principal									
Data de Criação de Contrato									
Utilizador que Criou Contrato									
Cliente Pleno									
Dia Débito Pleno									

Tabela 3-Informação das apólices

Para além da SPA_APOLICES existe um conjunto de tabelas que enriquecem a informação da apólice . Existe uma relação de 1:1 destas tabelas com a SPA_APOLICES uma vez que todas estas possuem o mesmo conjunto de Chave que as identifica (CODFAM + NUMINT).



FICHEIROS DE APÓLICES	
Agentes.....	MIG APO AGENTES
Apólices-dados complementares.....	MIG APO APOLICES
Atributos.....	MIG APO ATRIBUTOS
Bancassurance.....	MIG APO BANCASSURANCE
Beneficiários.....	MIG APO BENEFICIARIOS
Cartas Tipo.....	MIG APO CARTAS TIPO
Clausula Coseguro.....	MIG APO CLAUSULA COSEGURO
Coberturas.....	MIG APO COBERTURAS
Comissões Especiais.....	MIG APO COMISSOES ESPECIAIS
Comissões Especiais - Descontos.....	MIG APO COMISSOES ESPECIAIS DESCONTOS
Comissões Especiais - Mestre.....	MIG APO COMISSOES ESPECIAIS MESTRE
Conceitos Adicionais	MIG APO CONCEITOS ADICIONAIS
Coseguradores - Negócio Aceite.....	MIG APO COSEGUADORAS NEG ACEITE
Cosegurnet.....	MIG APO COSEGURNET
Credores Hipotecários.....	MIG APO CREDITORES HIPOTECARIOS
Distribuição.....	MIG APO DISTRIBUICAO
Documentos Digitais de Apólices	MIG APO DOC DIGITAIS
Encargos Fracionamento Prémio	MIG APO ENCARGOS FRAC PREMIO
Legendas de Apólices.....	MIG APO LEGENDAS APOLICES
Legendas de Produtos.....	MIG APO LEGENDAS PRODUTOS
Negócio Aceite.....	MIG APO NEGOCIO ACEITE
Negócio Aceite - Companhias.....	MIG APO NEGOCIO ACEITE COMPANHIAS
Negócio Aceite - Recibos.....	MIG APO NEGOCIO ACEITE RECIBOS
Observações.....	MIG APO OBSERVACOES
Outros Encargos de Fracionamento de Prémio.....	MIG APO OUTROS ENCARGOS FRAC PREMIO
Periodo Anuidade Apólices.....	MIG APO PERIODOS ANUIDADES APOLICES
Reservas Matemáticas Vida.....	MIG APO RESERVAS TOT V
Especiais.....	MIG APO TAXAS ESPECIAIS
Transações - Certificado.....	MIG APO TRANSACCAO CERTIFICADO
Transações.....	MIG APO TRANSACOES
Transferência de Mediação.....	MIG APO TRF MEDIACAO
EAD.....	EADHIS

Figura 27-tabelas de Apólices

A tabela SPA_Recibos contém a seguinte informação:

Apólices		Recibos				Vida
Dados Básicos		Transacção Financeira	Devoluções Documento Associado Transacção	Conta / Cliente / Identificador	Dados Adicionais	Dados Adicionais
Estorno	Não Estorno					
Nº Recibo	Nº Recibo	Nº Transacção	Nº Documento	Conta	Nº Lote	Nº Recibo
Nº Segurado	Nº Segurado	Montante	Documento	Nome Conta	Tipo Lote	Documento
Nome	Nome	Data Câmbio	Data Impressão Documento	Cliente	Fraccionamento	Data Emissão
Origem	Origem	Montante Liquido	Último Extracto	Nome Cliente		Data Limite
Documento	Documento	Transacção	Data Extracto	Referência		Montante
Produto	Produto	Estado Transacção		Identificador		Conta Cobrança
Nº Apólice	Nº Apólice	Regra Pag/Rec		Data Transacção		Produto
Tipo Recibo	Tipo Recibo	Juros		Lote Finalizado		Identificador
Período Validade	Fraccionamento	Plano Fraccionamento		Data Finalização		Situação
Data Emissão	Período Validade	Data Início		Tipo Finalização		Cliente
Utilizador Emissão	Data Emissão	Forma Reconciliação				Nome Cliente
Data Conferência	Data Dívida	Data Devido				
Utilizador Conferência	Data Limite Pag					
Data Autorização	Data Anulação Apólice					
Utilizador Autorização	Data Cobrança Recibo					
Data Cobrança Recibo	Canal Cobrança					
Valor a Pagar	Valor a Receber					
Canal Cobrança	Estado Documento					
Estado Documento	Data Última Alteração Doc					
Data Alteração Estado						

Tabela 4-Informação dos Recibos

Para além da SPA_RECIBOS , existem outras tabelas para complementarem os Recibos. Estas tabelas têm um conjunto de chaves em Comum : o CODFAM e o RECIBO (número de Recibo).

FICHEIROS DE RECIBOS	
Autorizações de Débito Direto.....	MIG_REC_AUT_DEBITO_DIRECTO
Autorizações de Débito Direto - Agentes.....	MIG_REC_AUT_DEBITO_DIRECTO_AGEN
Movimento Coseguro Cedido.....	MIG_REC_CEDIDO
Detalhe.....	MIG_REC_DETALHE
Distribuição.....	MIG_REC_DISTRIBUICAO
Guias Remessas.....	MIG_REC_GUIAS_REMESSAS
Movimentos SEPA.....	MIG_REC_MOV_SEPA
Movimentos.....	MIG_REC_MOVIMENTOS
Prémios Futuros.....	MIG_REC_PREMIOS_FUTUROS
Recibos-dados complementares.....	MIG_REC_RECIBOS
Recibos - Impostos.....	MIG_REC_RECIBOS_IMPOSTOS
Planos de Pagamento.....	MIG_REC_RECIBOS_PLANOS_PAGAMENTO
Referências Multibanco.....	MIG_REC_REF_MB
Faturas.....	MIG_REC_SAF_T_PT
SEPA Enviados.....	MIG_REC_SEPA_ENVIADOS

Figura 28-Ficheiro de Recibos

A tabela de Sinistros SPA_SINISTROS contém 3 chaves primárias, para além do CODFAM e o NUMINT o campo NUMSIN (número de Sinistro) também faz parte do conjunto das chaves da tabela.

Sinistros			
Ocorrência	Reclamante	Sinistro	Congénere
Nº Ocorrência	Reclamante	Nº Sinistro	Congénere
Situação	Reclamante Tomador	Linha Negócio	Nome Congénere
Data Ocorrência	Danos Materiais	Situação Sinistro	Nº Processo
Tipo Ocorrência	Danos Corporais	Objecto Seguro	Apólice
Descrição		Tipo Sinistro	Matrícula
Apólice		Cobertura	
Produto		Modo Regularização	
Tomador		Nome Gestor	
Documento Inicial		Data Abertura	
Gestor		Data Participação Sinistro	
Data Notificação		Data Encerramento	
Local Ocorrência		Data Reabertura	
Data Abertura		Motivo Reabertura	
Data Encerramento		Data Reencerramento	
Motivo Encerramento		Cosseguro	
Data Reabertura		Motivo Encerramento	
Danos Corporais		Motivo Reencerramento	
Data Reencerramento		Total Sinistro	
Danos Materiais		Data de Agravamento/Desconto	
Motivo Reencerramento		Ind de Multi Apolices	
Agravamento			
Data Agravamento			
Danos em outros veículos diferente de A e B			

Tabela 5-informação dos Sinistros

Para além da SPA_SINISTROS existem mais de 150 tabelas que complementam a informação do sinistro. As tabelas de Riscos foram divididas por tipo de Produtos fazendo um total de 200 tabelas. As Mestres é um outro tipo de tabelas que contém dados que não estão ligadas ao modelo. Todas estas tabelas mencionadas , foram criadas na base de dados STAGING_GP .Na STAGING_GP

Foram criados 3 schemas de tabelas distintos:

- O Schema GP (Generali Portugal)foi criado para as Tabelas SPAs da Staging GENSUDW .
- O Schema GPF (Generali Portugal Freeze) foi criado para as tabelas que estão identificadas para serem migradas na Migração Automática.
- As tabelas do Schema dbo foram criadas para controle interno .

5.1. ETLs DESENVOLVIDOS :

Para o carregamento das tabelas da STAGING_GP foram desenvolvidos por mim , vários ETLs com funções e objetivos distintos.

5.1.1. ETL GP_GetDataToStaging:

Após a fusão legal das duas empresas, a Generali decidiu que a venda de novos seguros a partir dessa data seriam vendidos através do sistema Tranquilidade e não entrassem no sistema Koolsite que mais tarde seria desligado.

De forma a disponibilizar ferramentas adequadas à rede comercial e back office durante este período de transição, foi implementado um front-end específico (SPA -Single Point of Access)- Ponto Único de Acesso com os dados básicos do cliente (apólices, recibos e sinistros) bem como informação da atividade comercial . Assim desta forma seria possível ter uma visão única dos clientes e os seus portfolios completos agregados das duas empresas.

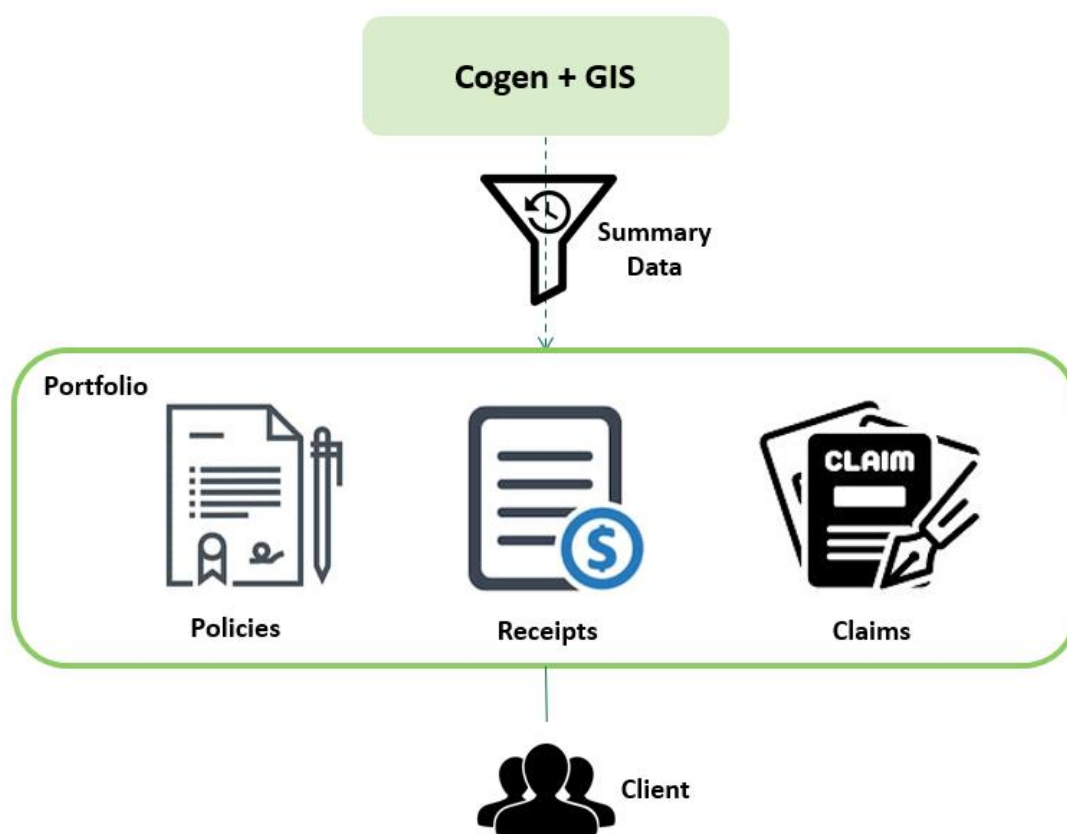


Figura 29-SPA Portfolio

A figura abaixo(figura 30) demonstra a agregação do Cliente da Generali que ainda não foi migrado para a Tranquilidade, foi feita uma pesquisa pelo NIF do cliente :

PESQUISA Nayara Mahomed Afzal

Cientes 2

Origem	Tranquilidade	NIF	Nome	Data Nasc.	Email	Apól.	Rec.	Ocorr.	WF
-	9120606735	123067197	AMERICO GONCALVES CUNHA	1953-03-18	-				
00737127	0500737127	123067197	AMERICO GONCALVES CUNHA	1953-03-18	-				

Figura 30-Exemplo de um cliente pesquisado no SPA

A primeira linha com a indicação a verde representa o Cliente da Tranquilidade que tem como número de cliente 9120606735.

A segunda linha a vermelho representa o cliente da Generali que tem como número 00737127 e que pós migração para o sistema tranquilidade passará a ser o número 0500737127.

O SPA foi construído a partir do ETL GP_GetDataToStaging .Este ETL não só faz o carregamento inicial dos dados do SPA ,mas também atualiza dinamicamente todas as alterações que são feitas no Koolsite. O Koolsite disponibilizou as tabelas na GENSUDW para a criação deste Ponto único de acesso e todas as tabelas foram criadas com o prefixo SPA. A equipa da Generali Portugal também criou uma tabela para o controle dos eventos a STG_OUT.SPA_CONTROLLO. Cada evento representa uma atualização a ser feita na nossa STAGING_GP.

Na STAGING_GP foram criadas as mesmas tabelas SPA usando o schema GP-Generali Portugal. Estas tabelas possuem os mesmos campos das tabelas SPA da GENSUDW. Para controlar as atualizações feitas na tabela foi adicionado mais dois campos às tabelas SPAs :

1. O STATUS_SYNC -campo para marcar o tipo de atualização que poderá apenas assumir 3 valores : o "I" que representa o Insert (novo registo), o "D" que representa o Delete (registo apagado) ou "U" que representa o Update (registo atualizado).
2. TIMESTAMP_SYNC -campo com o valor da data e hora que foi realizada uma atualização .

Como o SPA é a visão em Realtime dos portfólios dos negócios e este ETL corre todos os dias da semana de quinze em quinze minutos para estar atualizado conforme o KOOLSITE e GENSUDW. Para além de eventos de atualização o ETL também está preparado para receber eventos de Contabilidade, Conta efetivo e Zera Saldos.

A figura abaixo representa um Job do ETL com os 7 diferentes processos do ETL GP_GetDataToStaging:

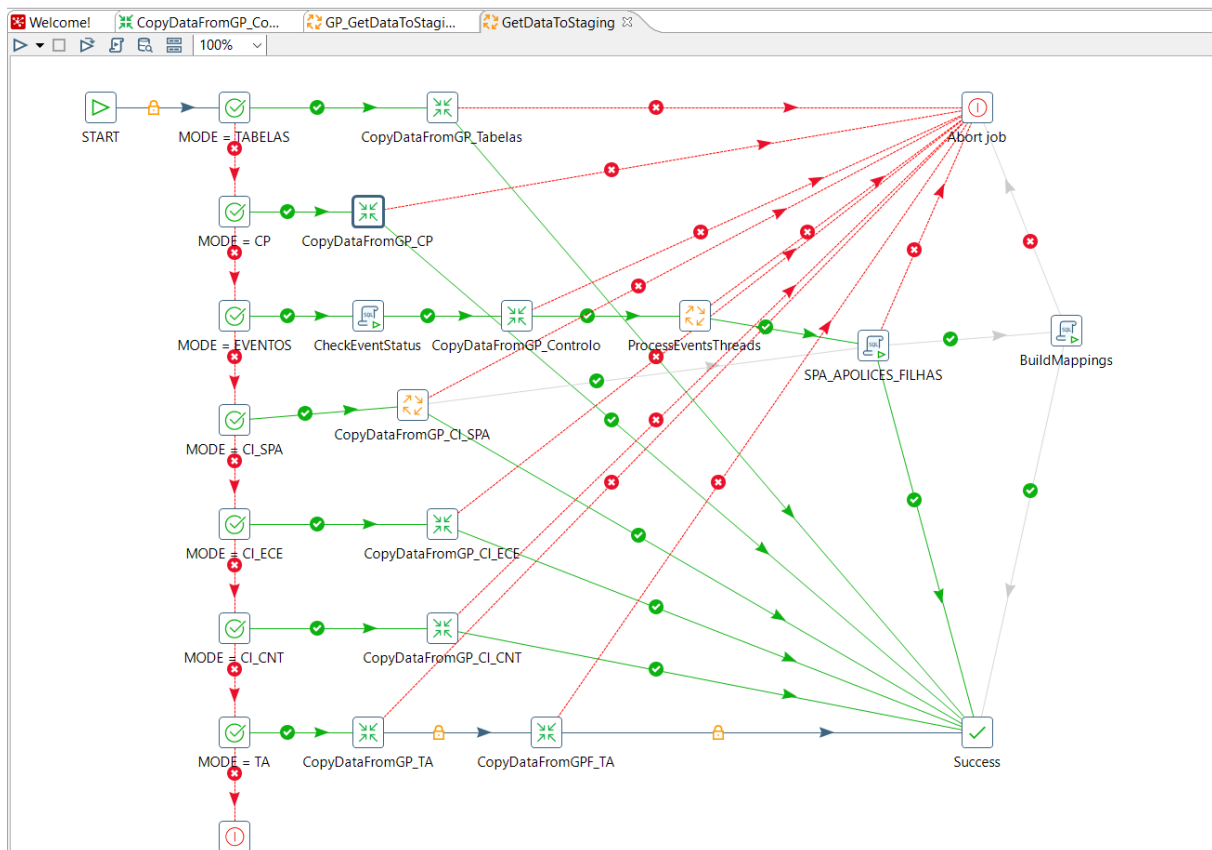


Figura 31-ETL GP_GetDataToStaging

1. O primeiro Processo MODE=TABELAS tem como objetivo o carregamento das tabelas Mestres .

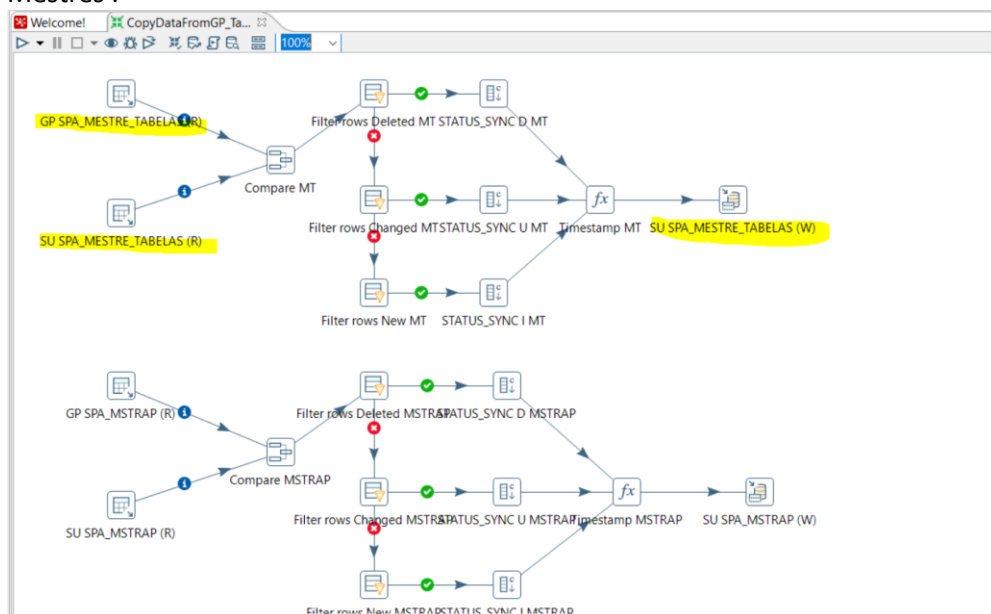


Figura 32-Carregamento das tabelas Mestres

Esta transformação começa com a leitura da tabela do lado da GP(Generali Portugal) ou seja GENSUDW, e ao mesmo tempo faz a leitura da mesma tabela do lado da SU(Seguradoras Unidas) STAGING_GP e a seguir compara os registos .Caso encontrar novos registos na GENSUDW que ainda não existam na STAGING_GP acrescentará na STAGING_GP , colocando o STATUS_SYNC a "I" de Insert. Caso algum registo deixe de existir da GENSUDW atualizará na STAGING_GP e colocará o

STATUS_SYNC desse registo a “D” de Deleted. E no caso de uma alteração nos dados a tabela da STAGING_GP será atualizada e o STATUS_SYNC desse registo passará a “U” de Updated.

2. O segundo Processo MODE=CP serve para o carregamento da tabela MESTRES de Códigos Postais:

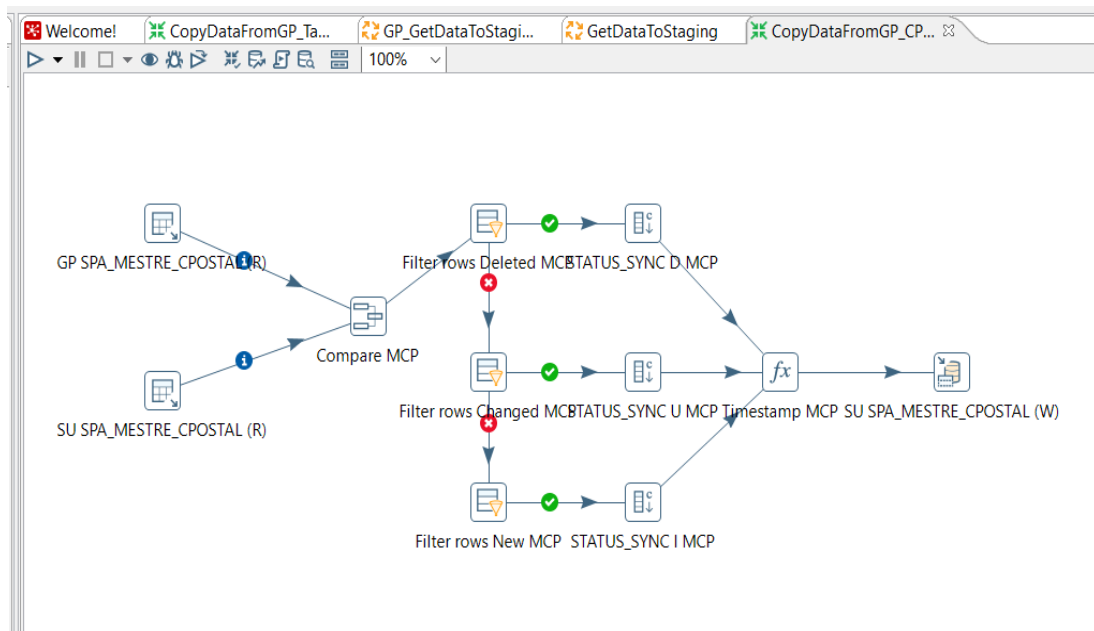


Figura 33-Carregamento dos Códigos Postais

3. O terceiro Processo MODE=EVENTOS é o Processo principal deste ETL. Este processo inicia com a leitura da tabela de eventos e escreve os novos eventos na tabela de controlo da SU.



Figura 34-carregamento da tabela de eventos

A seguir , podemos ter 10 jobs que vão buscar os eventos a serem tratados paralelamente pois há eventos com muitas alterações a serem feitas. Assim tendo 10 distintos JOBS a tratar paralelamente acelera o tratamento dos eventos :



Figura 35-Paralelismo de JOBs

Cada um desses 10 JOBS poderá tratar dos seguintes eventos: Atualização do SPA- CLIENTES; APÓLICES; RECIBOS; SINISTROS; RECEITAS; PROCESSO CONTA EFETIVO ; PROCESSO DE CONTABILIDADE; PROCESSO ZERA SALDOS :

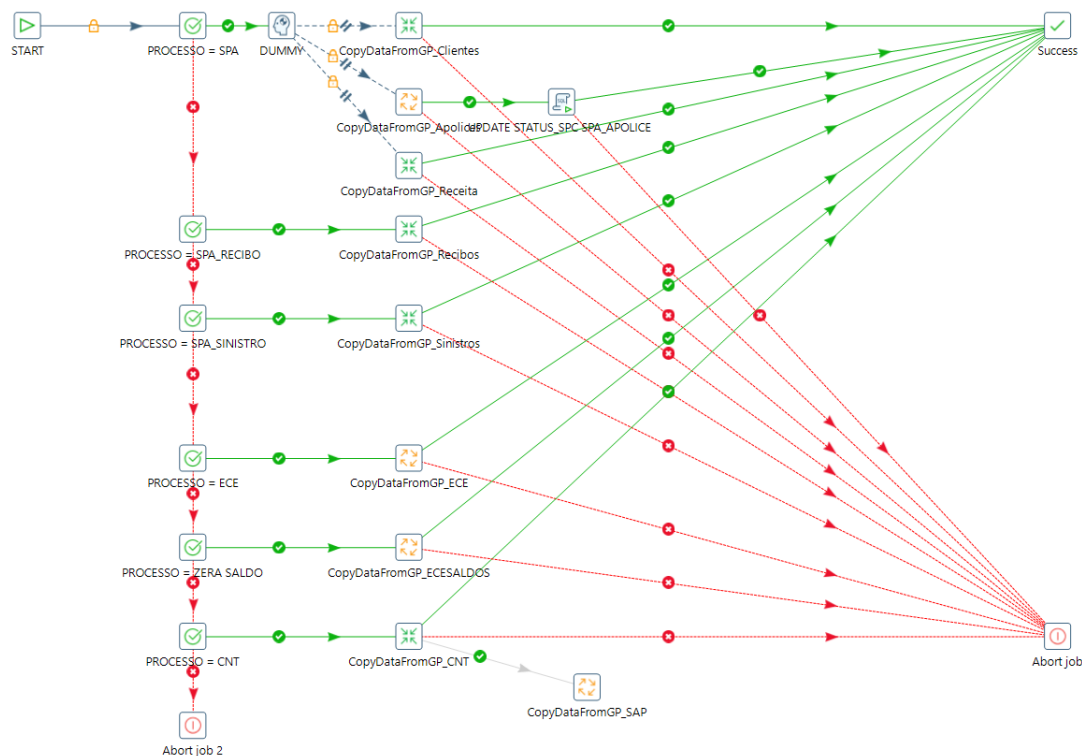


Figura 36-Tipos de Eventos

4. O quarto Processo tem como objetivo o carregamento completo de Todas as tabelas SPAs: CLIENTES;APÓLICES;RECIBOS;SINISTROS;RECEITA

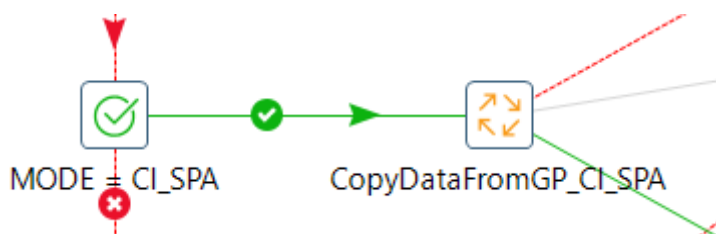


Figura 37-Carregamento Inicial do SPA

5. O quinto processo corresponde ao carregamento completo do Processo Conta Efetivo;
6. O sexto processo corresponde ao carregamento completo das tabelas de Contabilidade;
7. O sétimo processo corresponde ao carregamento das tabelas SPAs no ambiente de Testes de qualidade.

5.1.2. ETL GP_Staging_Controlo:

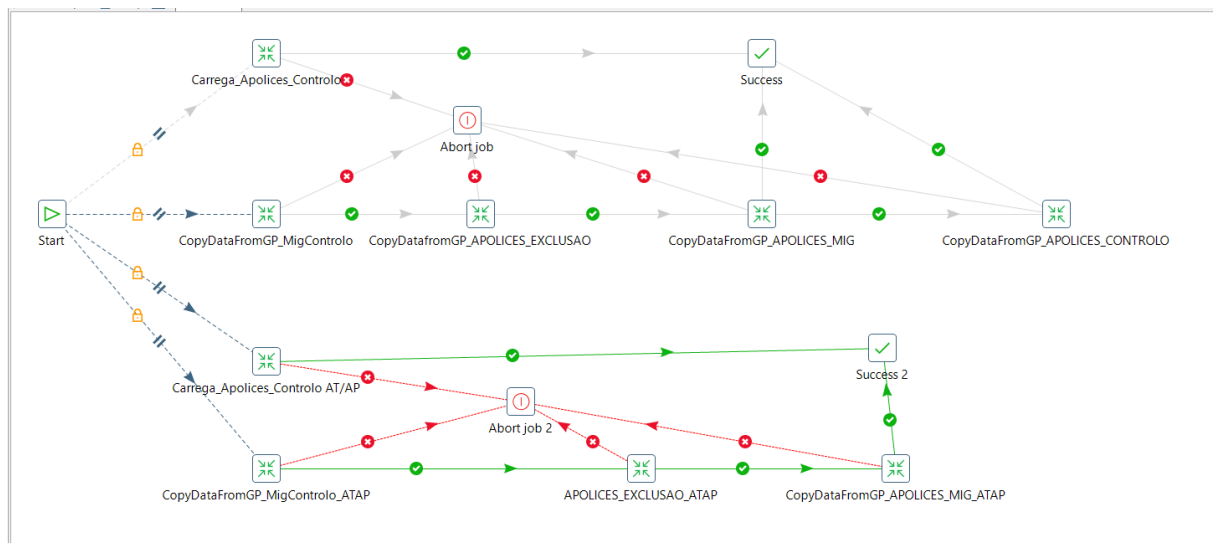


Figura 38-ETL GP_Staging_Controlo

Este ETL foi criado com objetivo de carregar as tabelas internas de controle para a Migração Automática. Como já foi referido anteriormente, na STAGING_GP foram criadas algumas tabelas com schema dbo para guardar as quantidades de apólices, recibos, sinistros que foram migrados em cada tranche.

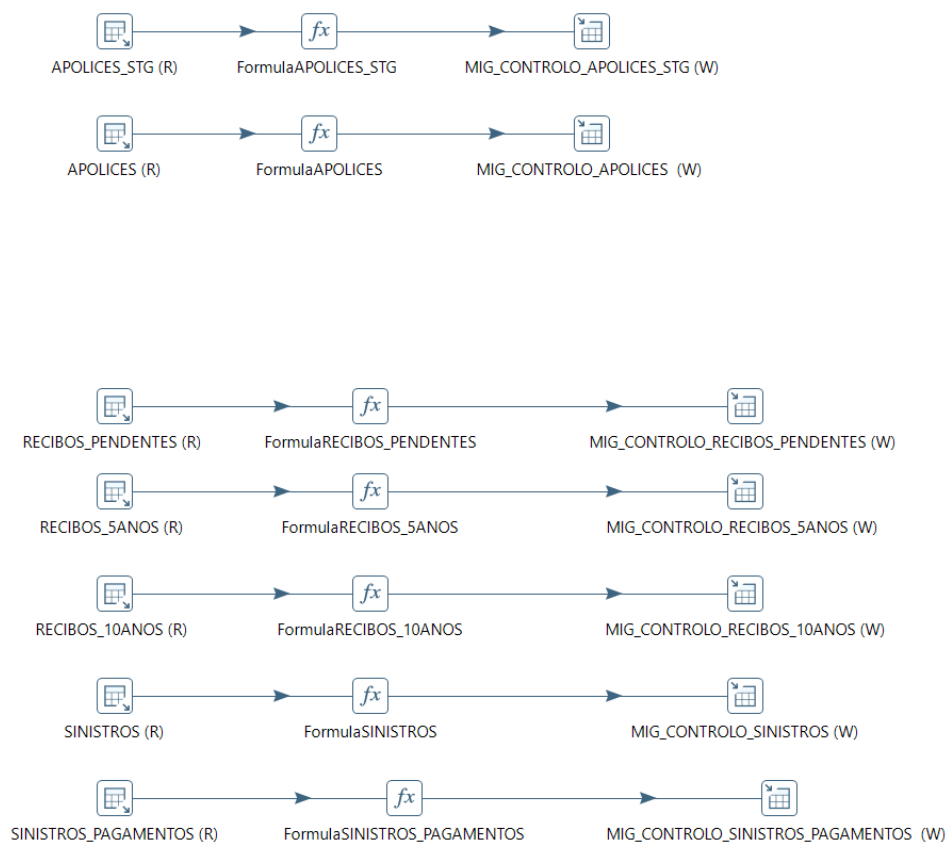


Figura 39-Carregamento das tabelas MIG CONTROLO

Neste ETL temos uma transformação para carregar uma tabela de controlo com as apólices que devem ser excluídas da Migração Automática de acordo com a tranche a ser migrada:

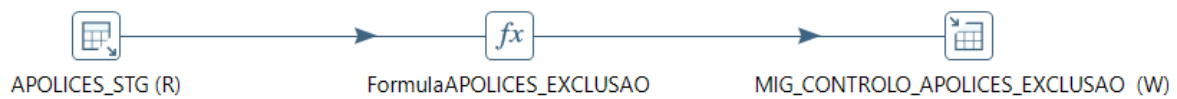


Figura 40-Carregamento da tabela APOLICES_EXCLUSAO

Para além do carregamento das tabelas internas de controlo , este ETL faz o carregamento da tabela GPF.APOLICES_CONTROLO conforme os produtos correspondentes de cada tranche . Depois de carregada esta tabela serve como base de extração das apólices na Migração automática.



Figura 41-Carregamento da tabela GPF.APOLICES_CONTROLO

5.1.3. ETL GP_GetDataToGPF:

O ETL GP_GetDataToGPF é o responsável pela Migração Automática.

Neste ETL as tabelas do schema GPF (Generali Portugal Freeze) da BD STAGING_GP são carregadas (Clientes, as Apólices ,os Recibos , os Sinistros e os Riscos) conforme a Tranche . O Modus Operandi da Migração Automática é o seguinte :

Automatic migration modus operandi

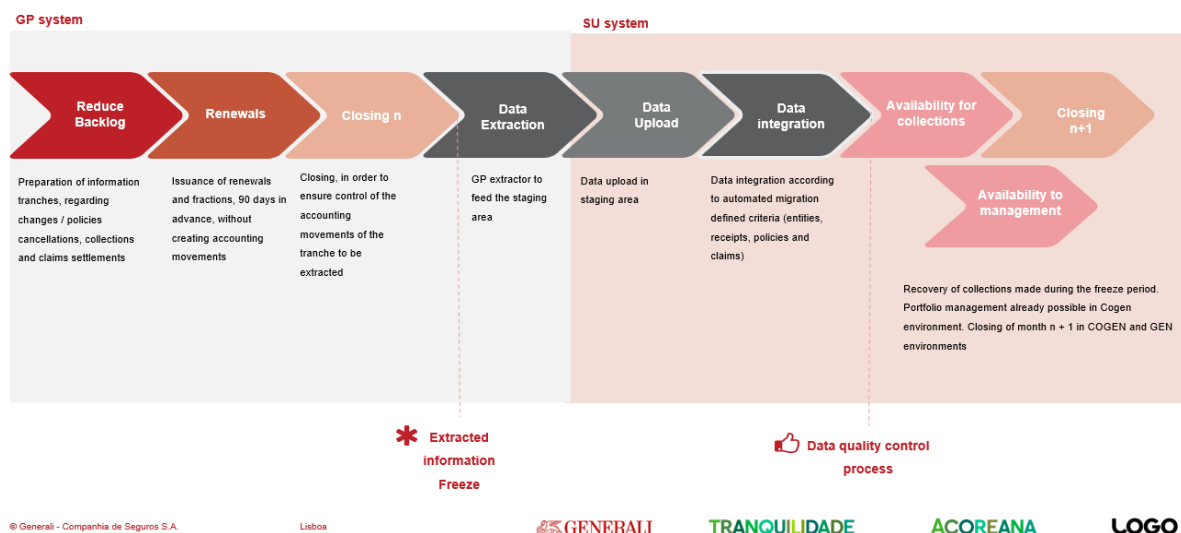


Figura 42-Modus Operandi da migração automática

Seguindo o esquema acima , a equipa GP faz uma extração completa das apólices da tranche a ser migrada. Assim que terminam o carregamento da BD GENSUDW , a nossa equipa SU inicia a execução do ETL GP_GetDataToGPF em Produção para o carregamento da STAGING_GP.

O Carregamento dos Clientes e das Apólices são processos independentes. Quando iniciei o estágio o processo de Clientes já tinha sido desenvolvido e as tabelas do schema GPF correspondente aos Clientes já estavam carregadas em Produção.

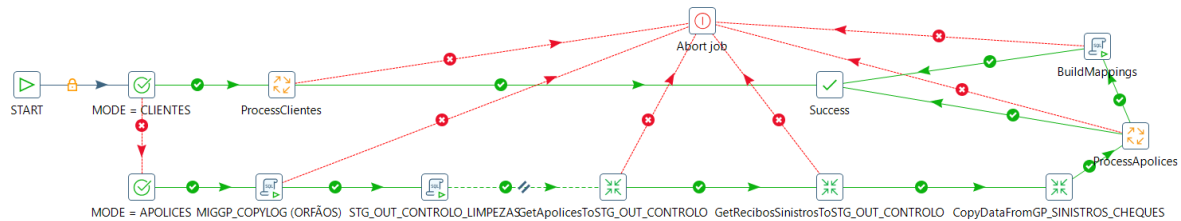


Figura 43-ETL GP_GetDataToGPF

Para o carregamento das apólices, o ETL começa por apagar todos os dados das tabelas de controlo das apólices, recibos e sinistros. Em seguida, a tabela de controlo das apólices é carregada com os dados da tabela GPF.APOLICES_CONTROLO e após esse carregamento as tabelas de controlo dos Recibos e Sinistros também são abastecidas. Após o carregamento das tabelas de controlo inicia-se o processo de cópia das tabelas. Para que a execução seja rápida foi criada 10 jobs que correm em paralelo para copiar os dados duma forma mais rápida.

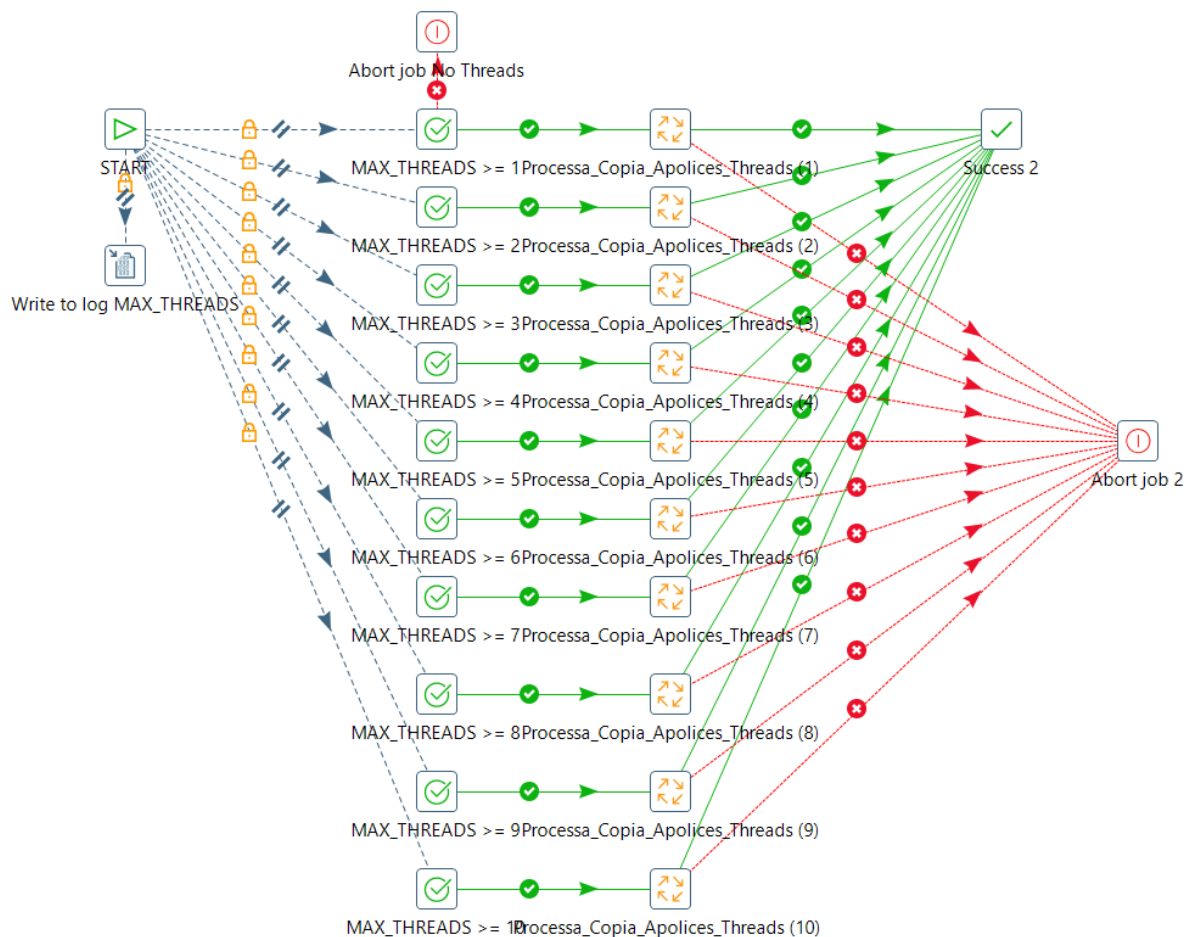


Figura 44-10 Threads para a cópia dos dados

Cada um desses 10 JOBS vai buscar um número de apólices para tratar:

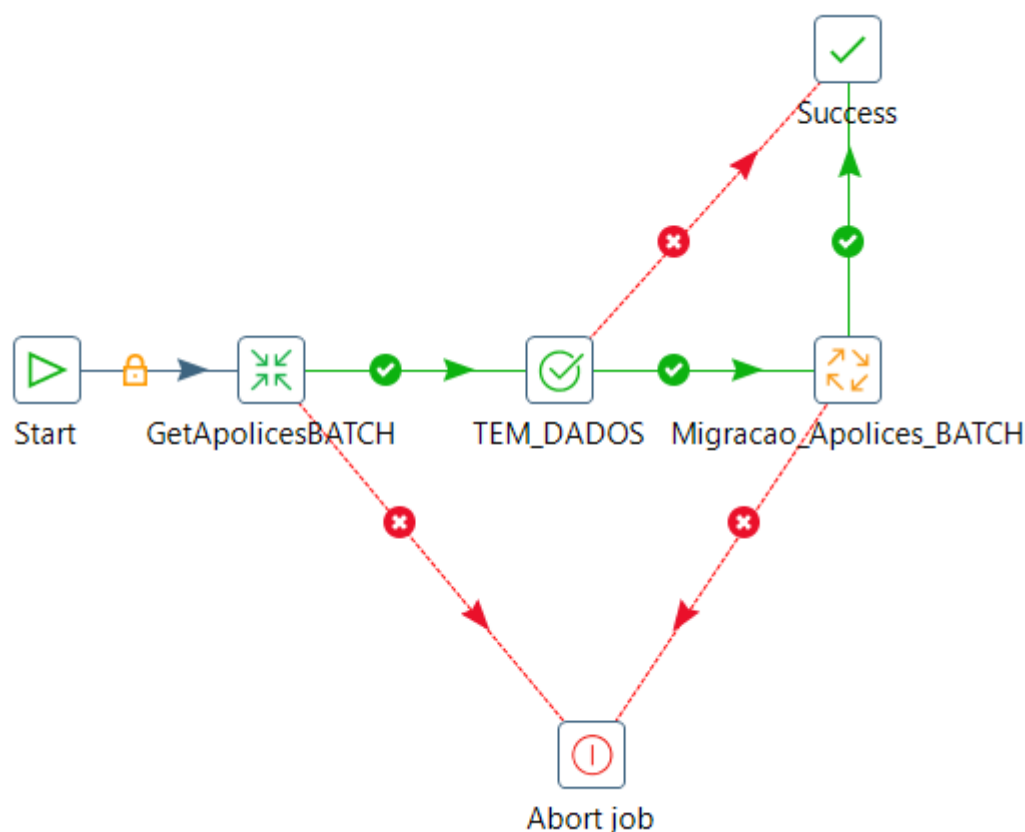


Figura 45-GetApolices

Normalmente executamos sempre 10 mil apólices por cada um desses jobs. A tabela abaixo mostra uma execução de testes feita para efeitos de controle do tempo total da cópia das tabelas para a STAGING_GP. Usou-se 1 milhão de apólices e a cópia levou cerca de 4 horas para ser executada:

Ambiente	PackageName	Duration	StartTime	EndTime	THREADS	Nº APOLICES P/THREAD	APOLICES PROCESSADAS
PR	GP_GetDataToGPF [APOLICES]	00:07:39	2021-02-21 17:35:04.997	2021-02-21 17:42:44.917	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:27:49	2021-02-21 17:05:04.047	2021-02-21 17:32:53.313	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:25:59	2021-02-21 16:35:04.487	2021-02-21 17:01:03.873	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:24:02	2021-02-21 16:10:04.66	2021-02-21 16:34:06.977	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:19:52	2021-02-21 15:50:04.053	2021-02-21 16:09:56.623	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:20:38	2021-02-21 15:25:04.42	2021-02-21 15:45:42.517	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:29:40	2021-02-21 14:55:04.603	2021-02-21 15:24:45.553	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:25:28	2021-02-21 14:25:04.867	2021-02-21 14:50:33.05	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:22:47	2021-02-21 14:00:04.427	2021-02-21 14:22:52.267	10	10 000	100 000
PR	GP_GetDataToGPF [APOLICES]	00:22:31	2021-02-21 13:35:04.893	2021-02-21 13:57:35.903	10	10 000	100 000
		3:46:25					1 000 000

Tabela 6-Tempo de execução de 1 milhão de apólices

A cópia é feita por apólice e por estrela .Se essa apólice tiver recibos, sinistros e riscos todos esses dados dessa apólice serão migrados. As chaves que fazem essa ligação da apólice com as outras tabelas são o CODFAM e o NUMINT que já foram mencionados em capítulos anteriores.

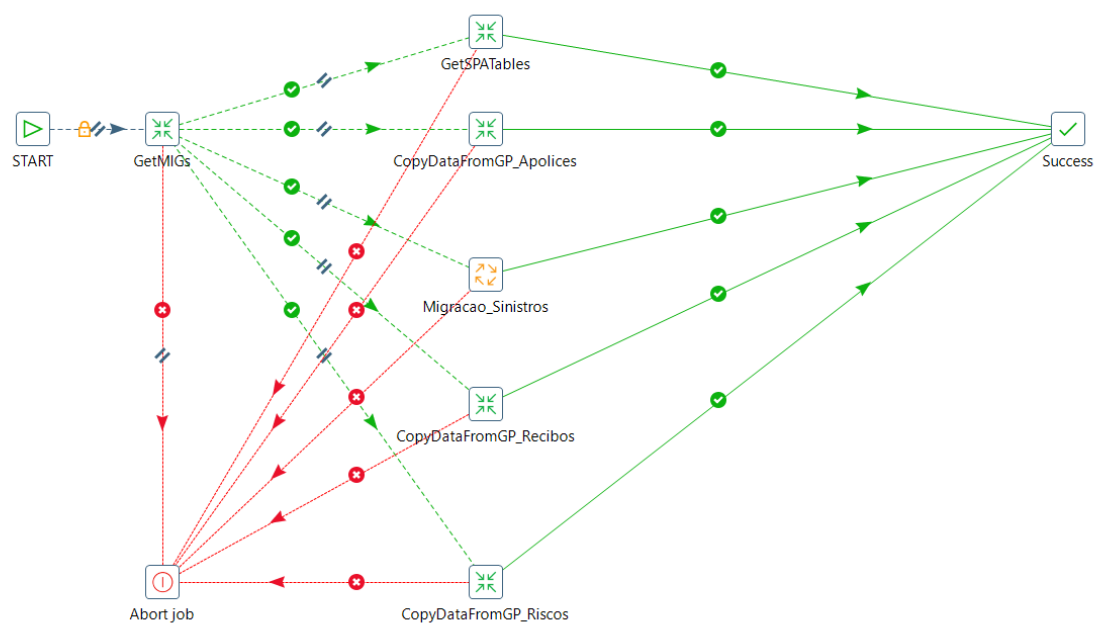


Figura 46-Migração por apólice

Primeiramente na é realizada a cópia dos dados das tabelas principais da Apólice, Sinistro e Recibos. A tabela de apólices é a SPA_APOLICES que contém todos os dados importantes de uma apólice. Os dados dos sinistros são armazenados na tabela MIG_SIN_SINISTROS e os de recibos na MIG_REC_RECIBOS.

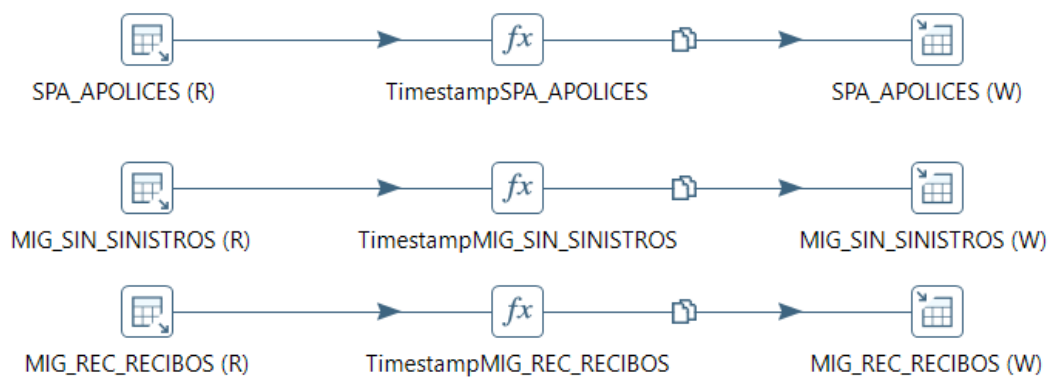


Figura 47-Cópia das tabelas principais

De seguida são feitas as cópias dos dados das outras tabelas das Apólices, Sinistros , Recibos e Riscos todas ao mesmo tempo em paralelo:

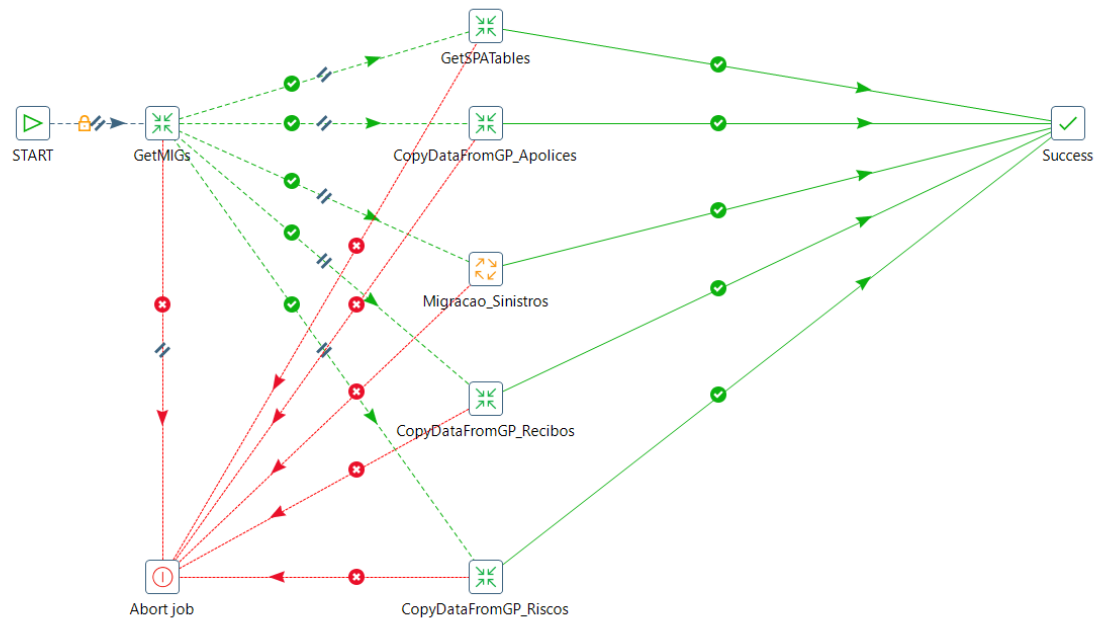


Figura 48-Migração das tabelas

A transformação responsável pela cópia das tabelas de apólices é a seguinte:

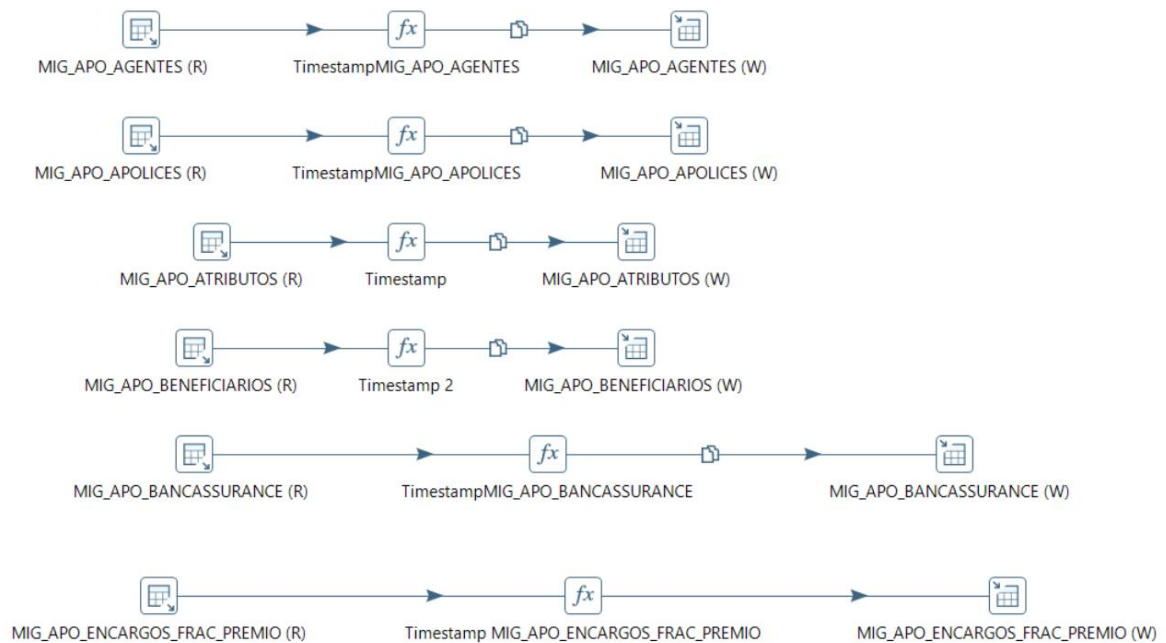


Figura 49-Cópia das tabelas de Apólices

A transformação que contém a cópia das tabelas de sinistros é a seguinte:

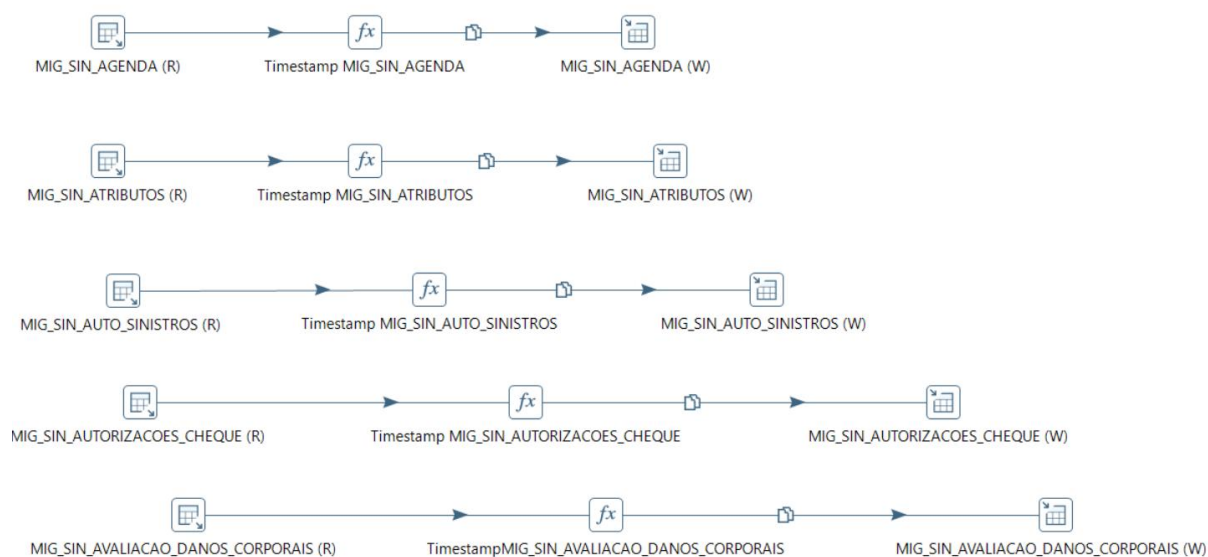


Figura 50-Cópia das tabelas de Sinistros

A transformação que contém a cópia das tabelas de Recibos é a seguinte:

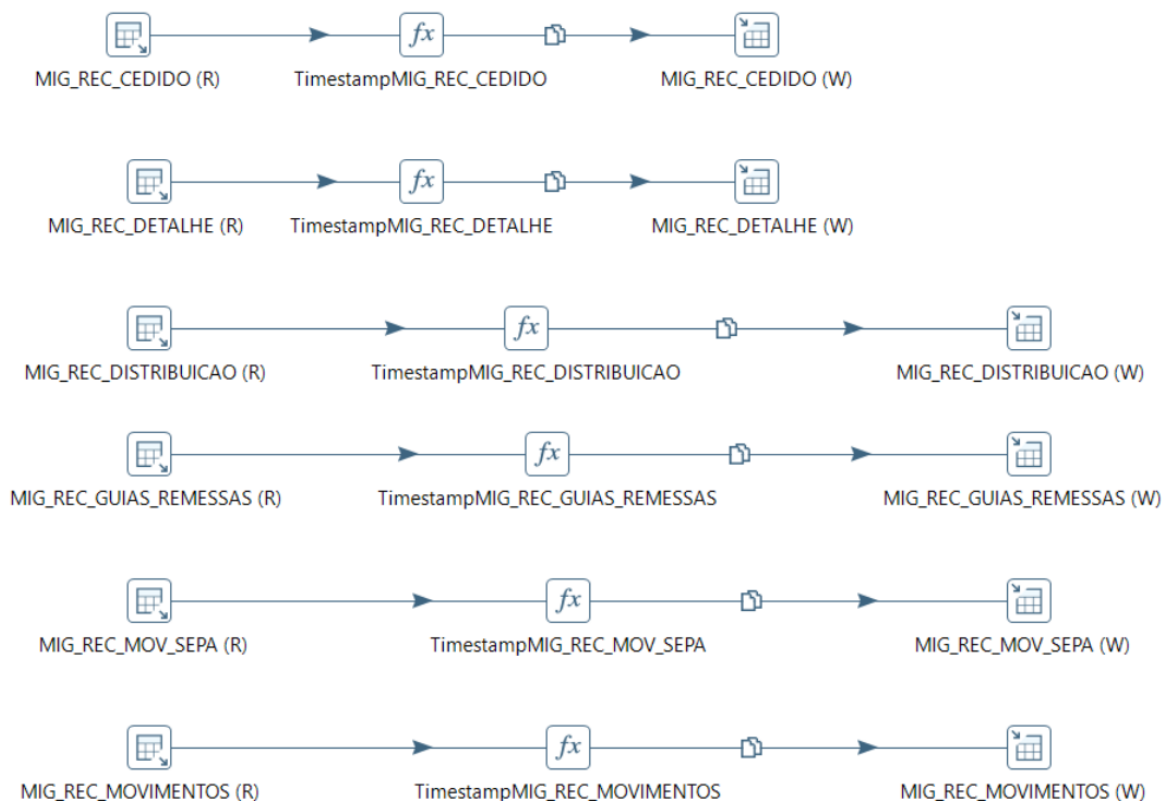


Figura 51-Cópia das tabelas de Recibos

A transformação que contém a cópia das tabelas de Riscos é a seguinte:

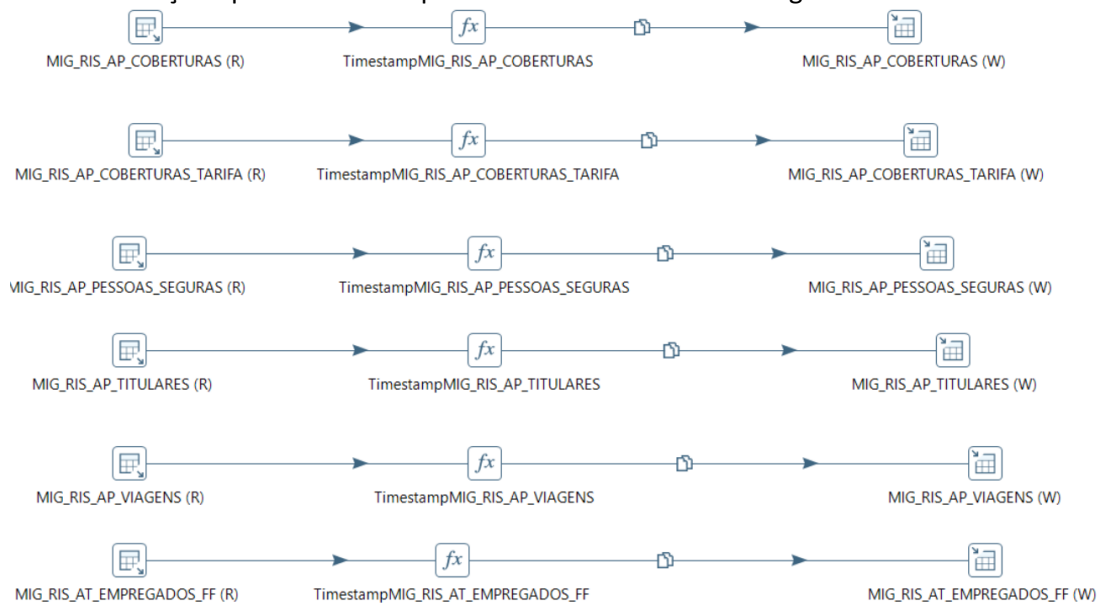


Figura 52-Cópia das tabelas de Riscos

Após terminar a execução deste ETL, a BD STAGING_GP fica completamente carregada e a migração para o Sistema Cogen poderá ser iniciada.

Caso haja algum erro durante a cópia dos dados, os dados são apagados e copiados novamente. Para a limpeza dos dados foram criadas Stored Procedures que apagam os dados da apólice em estrela. Durante a fase inicial, na primeira tranche foi implementada a Stored Procedure DeleteList que servia para apagar uma lista de apólices que ficaram em erro. Para a segunda tranche foi criada a Stored Procedure Delete Looping. Esta versão apaga as apólices que estão marcadas como erro ou em tratamento na tabela GPF.APOLICES_CONTROLO.

5.1.4. ETL GP_UPDATE_APOLICES_CONTROLO:

Este ETL foi criado para posicionar os Triggers da Migração Cogen assim que a BD STAGING_GP esteja carregada por completo. O Objetivo é marcar o STATUS_COGEN_POLICY- referente ao estado de migração cogen da apólice, STATUS_COGEN_NOTICE- referente ao estado de migração cogen do recibo e o STATUS_COGEN_CLAIM- referente ao estado de migração cogen do sinistro a:

- I- Início da Migração
- W-Trigger posicionado
- N-sem recibos/sinistros
- X-Finalizada Migração



Figura 53-ETL GP_UPDATE_APOLICES_CONTROLO

A figura acima(figura 53) representa as 3 principais transformações que servem de apoio para a Migração Cogen (cópia das tabelas da STAGING_GP para o sistema core da tranquilidade COGEN). A transformação UPDATE_COGEN_TO_I tem como objetivo marcar STATUS_COGEN_POLICY,STATUS_COGEN_NOTICE e o STATUS_COGEN_CLAIM a "I" (início migração) para os programas em Cobol migrarem as apólices ,os recibos e os sinistros para as tabelas internas da TRANQUILIDADE.

A função da transformação UPDATE_COGEN_CLAIM_W_TO_X é de marcar a X (finalizada migração) quando todos os sinistros de uma determinada apólice já estiverem migrados.

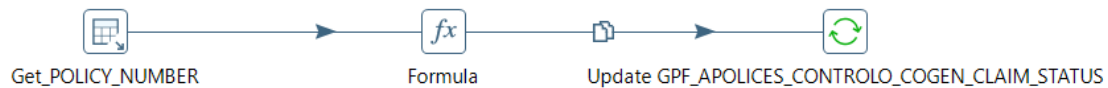


Figura 54-Transformação UPDATE_COGEN_CLAIM_W_TO_X

E a terceira transformação UPDATE_STATUS_LIBERTA_TO_X serve para marcar a finalização da migração cogen das apólices por completo(Apólice, recibos e sinistros):

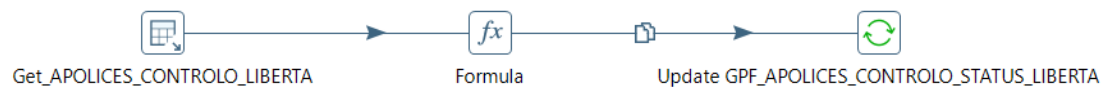


Figura 55-Transformação UPDATE_STATUS_LIBERTA_TO_X

6. CONCLUSÕES

Concluindo este projeto de estágio na Tranquilidade deu-me uma compreensão mais profunda da complexidade da arquitetura dum sistema ,a importância da qualidade de dados e acessibilidade, bem como a importância da estruturação conceitual para definir e alcançar os objetivos pretendidos. Durante a elaboração dos processos de ETL foi-me possível consolidar toda a formação académica do mestrado e aprofundar os meus conhecimentos na ferramenta Pentaho Data Integration . Fiquei bastante familiarizada com todas as tasks disponíveis no programa e aprendi quando e como é que as terei de usar, o que acabou por me permitir uma maior eficácia na execução do projeto.

No geral, o estágio proporcionou muitos benefícios a nível pessoal: integração numa instituição Externa, a capacidade de aprendizagem, a capacidade técnica, a autonomia, a capacidade de trabalho em equipa, iniciativa própria , visão abrangente, capacidade de comunicação (verbal e escrita) e, capacidade de organização e novas competências em soluções tecnológicas (SQL Server e PDI) conseguindo assim cumprir todas as metas pessoais e profissionais.

De uma maneira geral, este estágio constituiu uma excelente forma de exposição à realidade empresarial, através da minha integração como elemento produtivo, em projetos, com oportunidade de adquirir experiência nas áreas de conhecimento e ferramentas adotadas .

Fazendo uma autoavaliação pode-se afirmar que as tarefas realizadas durante o estágio foram positivas pois os projetos foram concluídos com sucesso e dentro dos prazos previstos. Este estágio permitiu a integração num ambiente empresarial, o qual era totalmente desconhecido, mas que se revelou bastante interessante tendo sido uma ótima experiência, tanto a nível pessoal como profissional.

7. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

O desenvolvimento do projeto teve algumas situações que dificultaram a normal progressão dos trabalhos a um ritmo constante. Primeiro, as principais dificuldades sentidas na realização do projeto surgem da utilização de algumas tecnologias, nomeadamente, SQL Server e Pentaho Data Integration, uma vez que no contexto académico é totalmente diferente do contexto empresarial.

Além da primeira limitação, o estágio foi realizado no âmbito de uma pandemia que resultou em esforços de trabalho remoto. O que exigiu muito esforço e vontade própria para aprender as ferramentas necessárias para a construção dos ETLs.

A recomendação clara para o trabalho futuro da Tranquilidade na continuação da Integração dos dados da Generali é para que nas próximas tranches seja definido com mais antecedência as apólices que serão migradas e as excluídas. E também que a equipa da Generali Portugal antes da migração automática disponibilize uma lista dessas apólices para facilitar nas contagens pós migração e no processo de controle de qualidade de dados.

8. BIBLIOGRAFIA

Damas, L. (2005). SQL: Structured Query Language(14). Lisboa: EDIÇÃO FCA – Editora de Informática, Lda. Obtido de <https://docplayer.com.br/56639763-14-a-edicao-edicao-atualizada.html>

Datapine, (2021). Get The Most Out of Smart Business Intelligence Reporting. Obtido de <https://www.datapine.com/blog/business-intelligence-reporting/>

Gartner. (2021). Magic Quadrant for Analytics and Business Intelligence Platforms. Obtido de <https://www.gartner.com/doc/reprints?id=1-1XYUYQ3I&ct=191219&st=sb>

Generali,(2021).O Grupo Generali. Obtido de <https://ww6.general.pt/generali/pt/quem-somos/o-grupo-general.html>

Greenplum Database, (2021). Overview of Greenplum Database Integrated Analytics. Obtido de <http://docs.greenplum.org/6-8/analytics/overview.html>

Group Kimball. (2013). Kimball Dimensional Modeling Techniques. Obtido de <http://www.kimballgroup.com/datawarehouse-business-intelligence-resources/kimball-coreconcepts/>

Gupta, V. R. (1997). An Introduction to Data Warehousing. Ed., 1997 Hitachi Vantara, (2021). Pentaho Data Integration. Obtido de https://help.pentaho.com/Documentation/7.1/0D0/Pentaho_Data_Integration

Inmon, W. H. (2005). Building the data warehouse (3). Nova Jersey, EUA: John Wiley & Sons, Inc.

Kimball R., Reeves L., Ross M E Thornthwate W. (1998). The Data Warehouse Lifecycle Toolkit - Expert

Methods for Designing, Developing, and Deploying Data Warehouses. John Wiley & Sons, Inc.

Kimball, R., Margy Ross (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3ed)

Larson, B. (2012). Delivering Business Intelligence with Microsoft SQL Server 2012 (3ed)

LINS, S. R.-p. (2003). Transferindo conhecimento tácito: uma abordagem construtivista. Rio de Janeiro: E-papers Serviços Editoriais.

Logianalytics ,(2021) Reporting in BI. Obtido de <https://www.logianalytics.com/resources/bi-encyclopedia/reporting-bi/>

Logi Report, (2021) The Goal of Reporting in BI. Obtido de <https://www.jinfonet.com/resources/bi-defined/bi-reporting/>

Malinowski & Zimányi (2009), Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications, Springer Berlin Heidelberg, Berlin.

Microsoft. (2021). Download SQL Server Data Tools (SSDT). Obtido de <https://docs.microsoft.com/en-us/sql/ssdt/download-sql-server-data-tools-ssdt?redirectedfrom=MSDN&view=sql-server-ver15>

MOSS, L. T., & ATRE, S. (2003). Business intelligence roadmap: the complete project lifecycle for decision-support applications. Boston: Pearson Education.

Oracle, (2021). Oracle SQL Developer is the Oracle Database IDE. Obtido de <https://www.oracle.com/tools/technologies/whatis-sql-developer.html>

REISSWITZ, F. (2012). Análise de sistemas: tecnologia e sistema da informação. Santa Catarina: Clube de Autores.

Tipos de Seguros .(2021).Obtido de <https://investidor.pt/tipos-de-seguros-ramo-vida-e-ramo-nao-vida/>

Santos, M.Y., Ramos, I. (2009) Business Intelligence - Tecnologias da Informação na Gestão de Conhecimento (2 ed.). Lisboa, Portugal: FCA.

Tranquilidade,(2021).A nossa história .Obtido de <https://www.tranquilidade.pt/tranquilidade/institucional/historia>

VERCELLIS, C. (2009). Business intelligence: data mining and optimization for decision making. Padstow: John Wiley & Sons.

9. ANEXOS

